GREAT BOOKS OF THE WESTERN WORLD

ROBERT MAYNARD HUTCHINS, *EDITOR IN CHIEF*

II.

EUCLID

ARCHIMEDES

APOLLONIUS OF PERGA

NICOMACHUS

GREAT BOOKS OF THE WESTERN WORLD

NOT TO BE TAKEN

28. GILBERT
    GALILEO
    HARVEY

29. CERVANTES

30. FRANCIS BACON

31. DESCARTES
    SPINOZA

32. MILTON

33. PASCAL

34. NEWTON
    HUYGENS

35. LOCKE
    BERKELEY
    HUME

36. SWIFT
    STERNE

37. FIELDING

38. MONTESQUIEU
    ROUSSEAU

39. ADAM SMITH

40. GIBBON I

41. GIBBON II

42. KANT

43. AMERICAN STATE
    PAPERS
    THE FEDERALIST
    J. S. MILL

44. BOSWELL

45. LAVOISIER
    FOURIER
    FARADAY

46. HEGEL

47. GOETHE

48. MELVILLE

49. DARWIN

50. MARX
    ENGELS

51. TOLSTOY

52. DOSTOEVSKY

53. WILLIAM JAMES

54. FREUD

ANGEL
ANIMAL
ARISTOCRACY
ART
ASTRONOMY
BEAUTY
BEING
CAUSE
CHANCE
CHANGE
CITIZEN
CONSTITUTION
COURAGE
CUSTOM AND CONVENTION
DEFINITION
DEMOCRACY
DESIRE
DIALECTIC
DUTY
EDUCATION
ELEMENT
EMOTION
ETERNITY
EVOLUTION
EXPERIENCE
FAMILY
FATE
FORM
GOD
GOOD AND EVIL
GOVERNMENT
HABIT
HAPPINESS
HISTORY
HONOR
HYPOTHESIS
IDEA
IMMORTALITY
INDUCTION
INFINITY
JUDGMENT
JUSTICE
KNOWLEDGE
LABOR
LANGUAGE
LAW
LIBERTY
LIFE AND DEATH
LOGIC
LOVE
MAN
MATHEMATICS

THE DIFFRACTION OF LIGHT, X-RAYS, AND MATERIAL PARTICLES

CHECKED

AN INTRODUCTORY TREATMENT

CHARLES F. MEYER

Associate Professor of Physics
University of Michigan

Checked
1987

THE UNIVERSITY OF CHICAGO PRESS
CHICAGO · ILLINOIS
1934

COPYRIGHT 1934 BY THE UNIVERSITY OF CHICAGO
ALL RIGHTS RESERVED. PUBLISHED SEPTEMBER 1934

COMPOSED AND PRINTED BY THE UNIVERSITY OF CHICAGO PRESS
CHICAGO, ILLINOIS, U.S.A.

Acc. No. 2144
Class No. B.9
Book No. [illegible]

C. THE FRAUNHOFER PATTERN DUE
(*Continuation of Chap. 5, Sec. 3*) . . .
D. THE ABERRATION OF THE CONCAVE GRATIN
Sec. 25) . . . . . .
E. THE ASTIGMATISM OF THE CONCAVE GRATING (*Con*
Sec. 27) . . . . . .
F. ROWLAND GHOSTS (*Continuation of Chap. 6, Sec. 37*) . .
G. THE DIFFRACTION OF X-RAYS (*Continuation of Chap. 8*) . .
H. THE REFRACTION OF ELECTRON WAVES (*Continuation of Chap. 9, and 10*) . . . . . rt"
"

INDEX

INDEX . . . . . .

. 323
. . . . 327
. 329
. 338
X-rays by Ruled Gratings 341
of X-rays by a Slit 346
rystal Spectrometer 348
orphous Solids, and Gases 359

CHAPTER 9. THE DIFFRACTION OF MATERIAL PARTICLES

avisson and Germer's Discovery 363
The Wave Theory of Matter 364

THE DIFFRACTION OF SLOWLY MOVING ELECTRONS
(Secs. 3–11)

3. Apparatus and Experimental Principles 367
4. The Nickel Crystal 370
5. Voltage, Colatitude, and Azimuth Curves 372
6. Plane-grating Beams 373
7. Gas Beams 374
8. Space-Lattice Beams 374
9. The Refraction of Electron Waves 377
10. The Bragg Reflection of Electron Waves 381
11. The Diffraction of Electrons by a Ruled Grating 383

THE DIFFRACTION OF FAST ELECTRONS
(Secs. 12–14)

12. Thomson's Method 383
13. The Diffraction of Electrons by Thin Crystals 386
14. The Diffraction of Fast Electrons by Thick Single Crystals . 391

15. The Scattering of Electrons by Gases and Liquids 393
16. The Polarization of Electron Waves 396

EXPERIMENTS WITH ATOMS, MOLECULES, AND POSITIVE IONS
(Secs. 17–21)

17. General Considerations and Technique 399
18. The Diffraction of Helium Atoms and Hydrogen Molecules . . 403
19. The Diffraction of Hydrogen Atoms 408
20. The Reflection or Diffraction of Metallic Atoms 410
21. Experiments with Positive Ions 412

APPENDIXES

A. The Method of Lunes and the Cornu Spiral (*Continuation of Chap. 3*) 415
B. The Pattern Due to a Single Straight Edge; Laboratory Directions (*Submitted in Connection with Chap. 4, Sec. 1*) 419

THE DIFFRACTION OF LIGHT

CHAPTER 1

GENERAL AND HISTORICAL

The word "diffraction" is derived from the Latin *dis*, meaning "apart" or "asunder," and *frangere*, "to break." The term "diffraction of light" signifies a certain breaking-up which a beam of light undergoes in passing an obstacle, and also signifies other types of breaking up which are fundamentally related to the one mentioned. The subject of diffraction covers a wide field.

1. Basic Facts and Early History.—The diffraction which occurs when a beam of light passes an obstacle was discovered by the Jesuit father Francesco Maria Grimaldi and was described by him in a book entitled *Physico-mathesis de lumine coloribus et iride.* This book was published two years after Grimaldi's death, by a friend—in the year 1665, at Bologna.

Grimaldi and other early investigators used the sun as their primary source of light. It is now often more convenient to use an artificial source. Referring to Figure 1, suppose that light from some primary source, S, such as an electric arc, passes through a condensing lens, L, and is focused upon an opaque plate with a tiny pinhole, O, in it. The pinhole acts as a secondary source, a *point* source from which the light diverges. Let us suppose that light from the hole passes an obstacle, D, the diffracting object, for example, a metal plate bordered by a straight edge perpendicular to the plane of the drawing, the edge being, let us say, vertical. If the shadow of the plate is received at P upon a sheet of white paper, one will observe that the boundary of the shadow is not sharp. Hence we must conclude that light is not propagated strictly in straight lines. The transition from illumination to darkness takes place gradually, though to be sure, in a small interval. Furthermore, outside of the shadow, parallel to its edge, will be found several bright and dark fringes. The distribution of light, considered as a whole, constitutes a certain *pattern.* Figure 2 shows, above, an enlargement of a photograph of the pattern, and, below, a diagrammatic representation of the distribution of light. The point E marks the edge of the

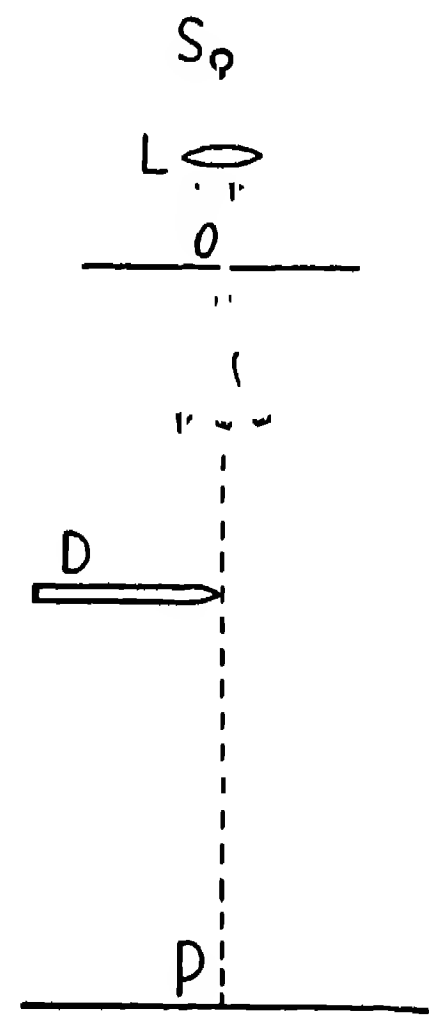


FIG. 1

so-called "geometrical shadow." The geometrical shadow is the area over which the shadow would extend if light traveled accurately in straight lines. The gradual transition from illumination to darkness which actually occurs, and the formation of the fringes, constitute the breaking-up of light which the word "diffraction" signifies in its original and primary meaning.

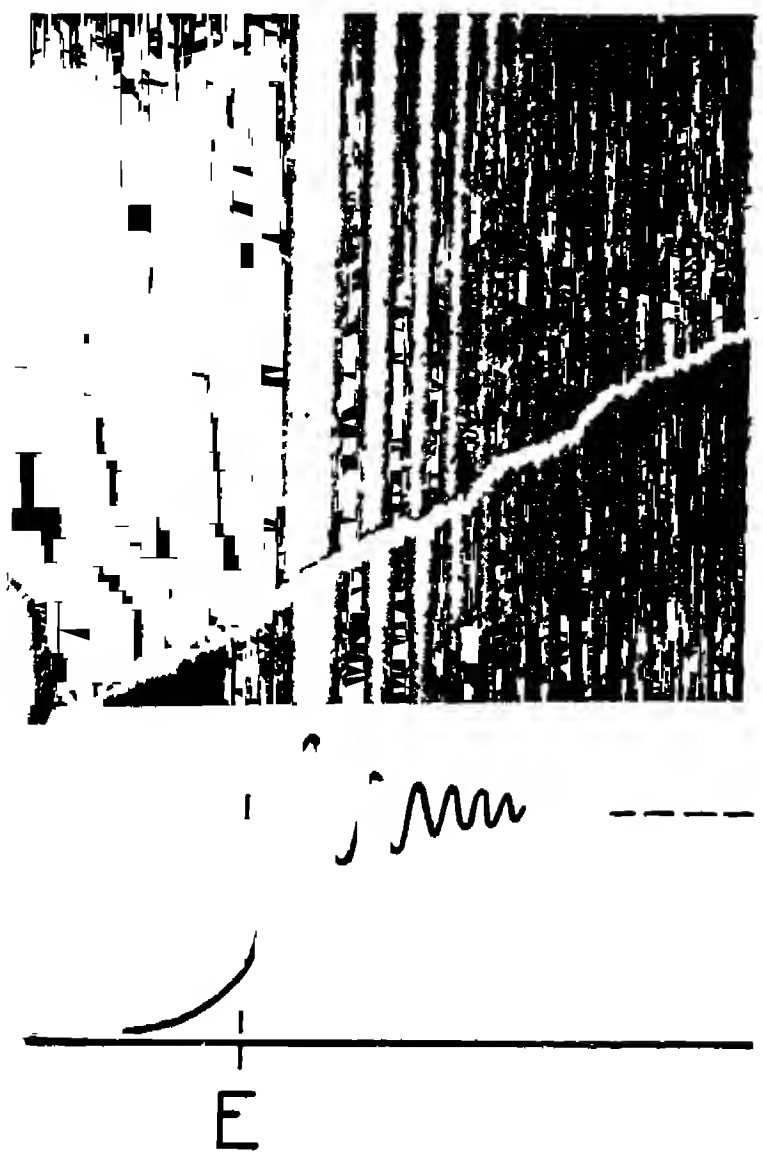


FIG. 2.—Single straight edge

The observations above indicated can be properly made only in a room which may be completely darkened.

The phenomena of diffraction show that the law of rectilinear propagation is not strictly true. Light is bent both into and out from the geometrical shadow. Yet we know that, in the large, the law of rectilinear propagation does hold. Hence we must explain why the law of rectilinear propagation holds approximately, and we must account quantitatively for the deviations from rectilinear propagation which occur. That is, we must account quantitatively for the observed diffraction patterns in all of those cases which lend themselves to analysis.

The problem of accounting for diffraction and that of accounting for rectilinear propagation are basically one and the same, but many years elapsed after the discovery of diffraction before this fact was realized.

A single straight edge gives rise to fringes only outside of the shadow, so-called "external" fringes. When these are to be observed visually, an intense primary source such as the sun or an arc light should be used. Even so, it is usually possible to see only the first three fringes. When white light is used, color effects are noticeable in the fringes. They are blue on the inside, the side toward the shadow, and red on the outside.

By way of further illustration, Figure 3 shows, above, an enlargement of the pattern due to a thin rod and below a diagrammatic representation of the pattern. There is now a set of external fringes on each side of the shadow and a set of "internal" fringes within the shadow. The edge of the shadow again lacks sharpness.

In general the secondary source of light must be a point source. But when the diffracting object presents only a straight edge or parallel straight edges, as in the cases thus far considered, it is permissible and usually ad-

vantageous to use a narrow slit, parallel to the edge or edges, as a source. But in no case may a source be used which has any considerable extension in area, because the patterns due to the various points of such a source would wipe each other out, owing to overlapping.

Figure 4 is a reproduction from Grimaldi's book and is a sketch of what he saw within the shadow of a narrow bar which at the end makes a right angle. A pinhole was used as the source. Grimaldi did not sketch, in this figure, the fringes which he saw outside of the shadow. Within the shadow he found two sets of fringes. One set runs, generally speaking, parallel to the bar, as we might expect from the previous illustration, but these fringes now round off at the corner and at the end of the bar. The other is a set of so-called "crested fringes" which as illustrated occur at the corner.

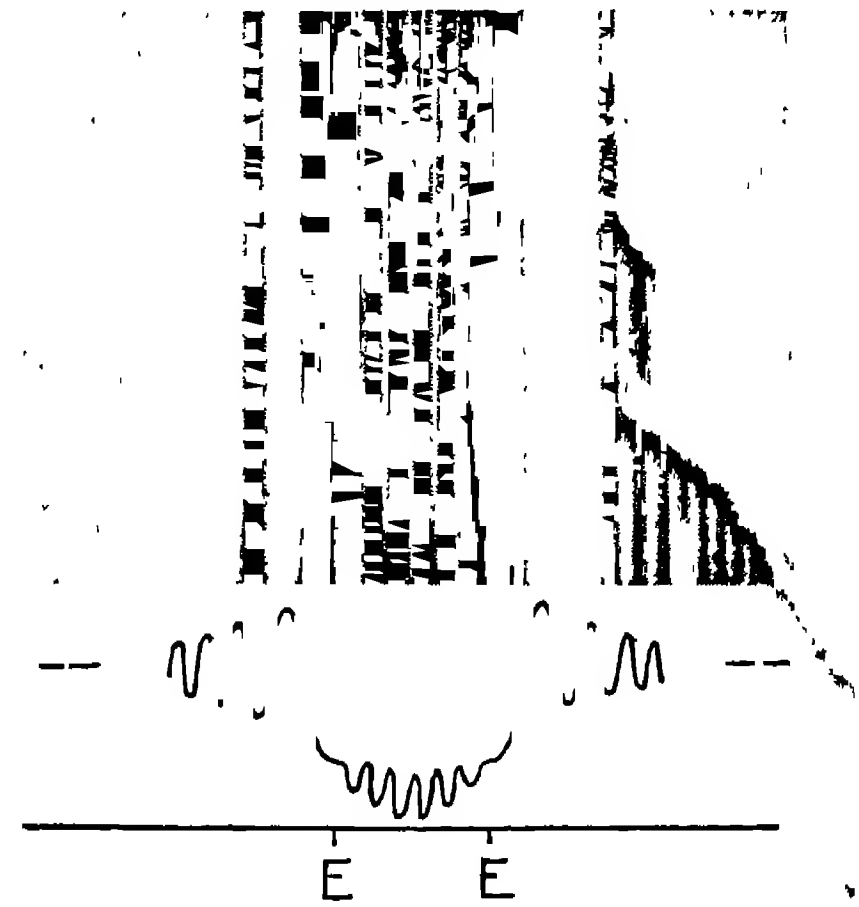


FIG. 3.—Pattern due to a thin rod, showing both external and internal fringes.

The internal fringes are much fainter than the external ones and hence require a much longer photographic exposure. The external fringes then become greatly overexposed. To obtain the foregoing reproduction two separate photographs were made, a relatively short one for the external fringes and one about five times as long for the internal fringes. The two photographs so obtained were afterward combined. Even so, some of the more delicate features of the pattern are bound to be lost in the process of photography and reproduction.

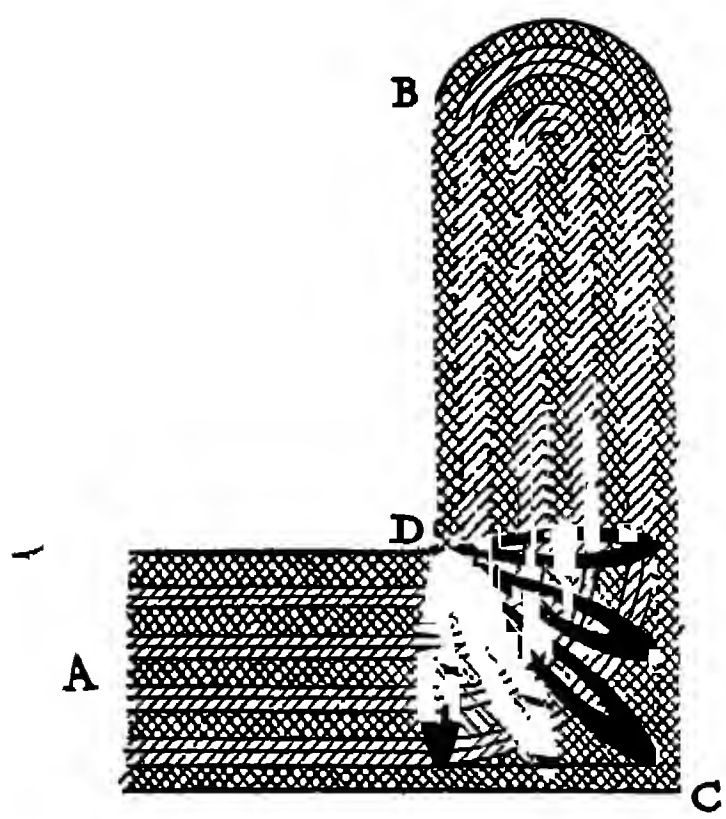


FIG. 4.—From Grimaldi

The diffraction pattern formed in a given instance is practically independent in its form, size, intensity, etc., of the material of which the object is made, so long as the material is opaque, and is practically independent also of whether the edge of the object be sharp or rounded, or whether the edge be a dull black and consequently absorbing or whether it be highly polished and reflecting. This circumstance was long not recognized; it was first pointed out by Fresnel in the year 1818.

Since the pattern is practically independent of the various above-mentioned factors, the cause of diffraction must be sought not in an effect of the

material of the diffracting edge upon light which passes close by, as was long supposed, but must be sought as being inherent in light in its own nature. The diffracting object acts primarily as a barrier, and diffraction theory must base its explanation upon the nature of light itself.

As has been said above, the material and nature of the edge are practically without effect, but they are not absolutely without effect. Ordinarily such influence as these factors exert are so slight that there is no observable difference in the pattern when the factors are varied at will. But by expressly using, for example, a polished metal bar of large diameter, which by reason of its large diameter presents a relatively flat surface, and then making observations close to the bar, we may observe effects which arise from light reflected by the bar. But as we withdraw from the bar these effects diminish rapidly, owing to the divergence of the reflected beam, and at only a reasonable distance back, even a large polished bar yields as nearly as can be observed the same pattern as an equivalent flat strip with razor-sharp edges.

Among the numerous interesting observations which Grimaldi made are that the shadow of a hair, at some distance behind the hair, is much broader than would be expected, and that when light passes through a small aperture, the cone beyond the aperture spreads out more than would be expected.*

(It has been claimed that Leonardo da Vinci discovered the diffraction of light about a century and a half before Grimaldi, but this claim is not well founded.)†

In 1672, seven years after the appearance of Grimaldi's book, Robert Hooke communicated to the Royal Society some observations on diffraction. He speaks of these as the discovery of a new property of light. Thus Hooke's statement implies that he had not heard of Grimaldi's discovery and that he, Hooke, discovered diffraction independently.‡

Before the close of the century Newton made some careful observations on diffraction, and in his *Opticks*, published early in the next century—in the year 1704 to be exact—he devoted the third and last book to diffraction, which he called "inflexion." This designation introduced by Newton long remained the common English term, but we have now returned to the use of Grimaldi's original term, "diffraction."

* For a more detailed account in English of Grimaldi's experiments see Joseph Priestly, *History of Vision Light and Colours* (London, 1772), pp. 171 ff., or, for an account in German, see Winkelmann's *Handbuch* (1st ed.) **2** (Part 1), 590.

† The claim in question was made by S. Libri in his *Histoire des sciences mathématiques en Italie*, **3**, 54 and 234, and is based upon an isolated observation of Leonardo's which while in itself interesting should not be called a discovery of diffraction. In any event the observation in no way influenced later developments.

‡ See Hooke, *Posthumous Works*, pp. 186–90, or Priestly, *loc. cit.*

Dedicated to

MARJORIE FLEMING MEYER

PREFACE

This book aims to provide a treatment of diffraction of a type for which the author has in his own experience long felt an urgent need, namely, a vivid presentation of the principal observational phenomena together with the theory which accounts for these, the development of the subject being carried reasonably far.

It is hoped that the book may prove of value to several classes of readers both as a medium of study and for reference:

Young physicists are here offered a sound basis for acquiring a knowledge of the subject, and mature physicists, familiar with the subject, are offered novel and improved exposition.

Workers in other sciences whose field requires knowledge of either the optics of light or the optics of X-rays should find at hand a treatment of diffraction not too mathematical, which they can study to advantage.

The substantive material of the early chapters of the present book and some of that of later chapters has been furnished to students intermittently for about ten years in the form of notes revised from time to time. In preparing these notes great effort was exerted toward making the treatment both graphic and thorough. As a result students were saved the serious perplexities with which they are ordinarily confronted when they enter upon the study of diffraction. The notes have again been carefully revised and new material has been added. The last two chapters of the present book constitute addition in their entirety. The author plans to use appropriate portions of the now printed book as a text in that part of his course in optics which deals with diffraction. It is hoped that other instructors may carefully consider the advisability of similar or somewhat similar use.

The author wishes to thank various of his colleagues of the University of Michigan for reviewing those chapters of the manuscript which deal with the diffraction of light. In addition he wishes to thank Professors G. A. Lindsay and J. M. Cork of this university, Professor S. K. Allison of the University of Chicago, and Dr. J. M. Robertson of the Royal Institution, for making suggestions regarding the chapter on the diffraction of X-rays, also Drs. C. J. Davisson and L. H. Germer of the Bell Telephone Laboratories for the same kindness in connection with the chapter on material particles. Finally, he wishes to thank Drs. J. A. Anderson and H. D. Babcock of the Mount Wilson Observatory and Professor J. E. Mack of the University of Wisconsin for reviewing the sections dealing with the concave grating and with error of ruling.

C. F. M.

Ann Arbor, Michigan
April, 1934

TABLE OF CONTENTS

THE DIFFRACTION OF LIGHT

(Chaps. 1–7)

CHAPTER 1. GENERAL AND HISTORICAL

1. Basic Facts and Early History 1
2. The Views of Newton 5
3. The Views of Young 7
4. Fresnel's Investigations 10
5. The Principles of Fresnel's Theory 14
6. Sketch of the Life of Fresnel 16
7. Development of the Wave Theory of Light 17
8. Elastic Solid Theory of the Ether 21
9. Advent of the Electromagnetic Theory 22
10. Recent General Developments 25
11. Developments in Diffraction Theory 26

CHAPTER 2. THE METHOD OF ZONES

1. Fresnel Zones 28
2. Retardation, or Path Increment 29
3. Dimensions of Zones of Low Number 30
4. Circular Aperture; Small Number of Zones 31
5. Zone Plates 32
6. Equations of Wave Motion 33
7. The Contribution from an Elementary Area 35
8. The Summation of Simple Harmonic Disturbances 36
9. Elementary Zones 38
10. The Inclination Factor 40
11. Circular Aperture; Any Number of Zones 42
12. The Complete Wave Front 43
13. The Areas of the Zones 44
14. Circular Obstacle 45
15. Rectilinear Propagation 47
16. Irregular Obstacle and Circular Aperture 51

CHAPTER 3. THE METHOD OF LUNES AND THE CORNU SPIRAL

1. Half-Period Lunes 53
2. Elementary Lunes 54
3. Vibration Polygon Based upon Elementary Lunes 55
4. The Geometry of the Cornu Spiral 56
5. The Cornu Spiral as a Vibration Curve 60

CHAPTER 4. TYPICAL DIFFRACTION PATTERNS OF THE FRESNEL CLASS

1. The Straight Edge 66
2. The Single Slit 70
3. The Bar 73
4. The Double Slit 76
5. The Circular Aperture 77
6. The Pinhole Camera 80
7. The Circular Obstacle 80
8. The Square Corner; the Tapered Slit 82
9. Graphical Methods for Circular Aperture and Obstacle . . . 83

CHAPTER 5. THE FRAUNHOFER CLASS OF DIFFRACTION PHENOMENA

1. Optical System 85
2. Rectangular Aperture 86
3. The Fraunhofer Pattern Due to a Circular Aperture 90
4. General Principles of Fraunhofer Diffraction 92
5. The Fraunhofer Pattern for a Double Slit 94
6. Rectangular Aperture (*Concluded*) 96
7. Distinctions between Fraunhofer and Fresnel Diffraction . . 101
8. Virtual Diffraction Patterns 104
9. The Symmetry Properties of Fraunhofer Patterns 106
10. The Theorems of Bridge 107
11. The Life and Work of Joseph von Fraunhofer 108

CHAPTER 6. THE DIFFRACTION GRATING

1. Diffraction by Narrow Equidistant Slits 112
2. Formation of Spectrum 113
3. Overlapping Spectra 115
4. The Grating Law 115
5. Gratings Having a Small Number of Lines 116
6. Gratings Having Many Lines 120
7. Periodic Structure as Prime Requisite 124
8. The Reflection Grating 125
9. Depth and Form of Grooves 125
10. The Ruled Grating; Historical 126
11. Rowland's Ruling Machines 128
12. Replicas 131
13. Crossed Gratings 132
14. Wire Gratings; Gratings Ruled through a Silver Film 132
15. Intensity Ratios; Absent Spectra 134
16. Polarization 136
17. The Laminary Grating 137

THE PLANE GRATING IN PRACTICE

(Secs. 18–22)

18. Direct Vision 138
19. The Common Spectrometer 139
20. The Spectrograph 141
21. The Infra-red Spectrometer 142
22. The Echelette 144

THEORY OF THE CONCAVE GRATING

(Secs. 23–27)

23. General Properties 146
24. Theory of Focusing along Rowland Circle 149
25. Some Numerical Evaluations 152
26. General Focal Relation 155
27. Astigmatism 157

CONCAVE-GRATING MOUNTINGS

(Secs. 28–33)

28. The Paschen Mounting 161
29. The Abney Mounting 162
30. The Rowland Mounting 163
31. The Eagle Mounting 164
32. The Mounting in Parallel Light 165
33. Mountings for the Extreme Ultra-violet 167

ERROR OF RULING

(Secs. 34–43)

34. General 170
35. Curvature and Lack of Parallelism of the Grooves 171
36. Periodic Error of Spacing 172
37. Rowland Ghosts 174
38. Lyman Ghosts 180
39. Ghosts Very Close to the Parent-Line 185
40. Error of Run 185
41. Accidental Error 190
42. Abrupt Error 191
43. Variation of Groove Form 192

CHAPTER 7. SUNDRY RAMIFICATIONS OF DIFFRACTION THEORY

1. Permitted Aberration 193
2. Resolving Power of the Telescope 194
3. Resolving Power of the Grating 197
4. The Law of Extreme Path 199
5. Resolving Power of Prism 202

6. Resolving Power of Microscope 205
7. Abbe's Theory of Microscopic Vision 212
8. Measurement of Star Diameters 216
9. Measurement of Diameters of Microscopic Particles 221
10. Luminosity of the Diffracting Edge (Observation) 222
11. Luminosity of the Diffracting Edge (Theory) 225
12. Babinet's Principle 233
13. The Diffraction of Light by Small Particles 237
14. Young's Eriometer 238
15. Meteorological Phenomena 240
16. Quetelet's Rings 244
17. Rayleigh Scattering 244
18. Diffraction by a Slightly Transparent Screen 251
19. Diffraction by a Thin Transparent Lamina 252
20. Talbot's Bands 253
21. The Colors of Mixed Plates 253
22. The Action of a Grating on Single Pulses 255
23. Group Velocity 256
24. The Action of a Prism on Single Pulses 259
25. Critique of Fresnel's Theory of Diffraction 264
26. The Dynamical Theory of Diffraction; General 269
27. The Theory of Kirchhoff 271
28. Remarks concerning the Theory of Sommerfeld 274
29. Diffraction at Large Angles 275
30. Observation Close to the Diffracting Edge 276
31. Diffraction and the Quantum Theory 277

CHAPTER 8. THE DIFFRACTION OF X-RAYS

1. Roentgen's Discovery 279
2. The Nature of X-rays 280
3. Laue's Discovery 286
4. Concerning Crystal Structure 288
5. The Bragg Theory 292
6. Some General Aspects of Space-Lattice Diffraction 296
7. Spectrographs for Soft X-rays 300
8. Volume Effect; Spectrographs for Hard X-rays 306
9. Methods of Crystal Analysis 310
10. Rotation Photographs 311
11. The Powder Spectrograph 313
12. The Laue Theory (Introductory) 315
13. Diffraction by a Row of Particles 316
14. Diffraction by a Plane Net 318
15. Crossed Gratings 319
16. Diffraction by a Row of Particles (*Concluded*) 320
17. The Space Lattice 321

The term "inflection" is, however, still used. It is today used to designate specifically the bending *inward* of light by the obstacle, and the term "deflection" is used to designate the bending *outward* of light by the obstacle. When light passes an obstacle some of it is inflected and some of it is deflected.

In the century 1700–1800, diffraction was investigated with keen interest by Delisle, Mairan, Maraldi, and numerous others.*

Observation accumulated for well over a hundred years, but during all of this time no real explanation of the observed phenomena was advanced. During this period diffraction was a thing apart—an appendage to what then constituted the subject of optics.

This situation was destined, however, to change radically. In the period roughly between 1800 and 1825, great progress was made in our understanding of the nature of light, principally through the work of Young and Fresnel. Moreover, one of the important steps in this progress, perhaps the most important, was the divination by Fresnel of the cause of diffraction and, at the same stroke, of the cause of approximate rectilinear propagation. Thenceforth the subject of diffraction has occupied a place of prominence in physical science, this prominence increasing gradually as subsequent developments continued to extend the application of diffraction theory far beyond the original bounds. Today this theory, followed into all of the matters to which it has been found to have application, takes us far afield indeed. It takes us into many subjects upon which, offhand, one would not surmise that it had any bearing whatsoever.

We shall now consider in succession the two principal views of the cause of diffraction which were held before the day of Fresnel. By considering these views we shall gain historical perspective and deepen our insight into the theory which lies ahead of us.

2. The Views of Newton.—Because waves in general bend around obstacles to a much greater degree than light is observed to bend around them, Newton rejected the theory of his contemporary, Huygens, that light is due to waves in an all-pervading ether, and propounded the view that light is

* The following list of references is taken from Winkelmann, *op. cit.* (1st ed.), **2,** Part I, 593: Delisle, *Mem. Acad. Roy.* (Paris, 1715), pp. 147, 166; Maraldi, *ibid.* (Paris, 1723), p. 111; Mairan, *ibid.* (Paris, 1738), p. 53; s'Gravesande, *Physica elementa* (3d ed., 1742), p. 725; Le Cat, *Traité des sens* (1740), p. 299; Dutour, *Mem. Sav. Etr.*, **5,** 635; **6,** 19 and 36, 1768–74; *Jour. de phys.* (Rozier), **5,** 120, 230, 1775; **6,** 135, 412; Du Sejour, *Mem. Acad. Roy.* (Paris, 1775), p. 265; Marat, *Découvertes sur la lumière* (Londres, 1780); Hopkinson and Rittenhouse, *Trans. Amer. Phil. Soc.*, **2,** 201, 1786; Comparetti, *Observations opticae de luce inflexa Patav.* (1787); Stratico, *Saggi di Pad.*, **2,** 185, 1789; Brougham, *Phil. Trans.* (1796), p. 227; *ibid.* (1797), p. 352; Jordan, *The Observations of Newton* (London, 1799); *Gilb. Ann.*, **18,** 1. To this list should be added Delisle, *Memoirs* (St. Petersburg, 1728), p. 205.

due to minute corpuscles emitted by luminous bodies. Newton believed in the existence of an ether, but this ether in his mind played only a secondary rôle in the propagation of light. Condensations and rarefactions in the ether supposedly influenced the behavior of the corpuscles, especially at the bounding surface between two mediums such as air and glass.

Newton's work on diffraction was of a fragmentary nature. He made a number of interesting observations but was interrupted and never found an opportunity to resume them. His expressed views on the cause of diffraction are little more than offhand conjectures. They are by no means all in evident harmony. Newton regarded the diffracting edge as exerting a force of attraction upon corpuscles which pass extremely close by, causing the paths of these corpuscles to be turned inward, thus accounting for the inflected light. Referring to Figure 5 to illustrate Newton's views, let us suppose the light is coming from above in the figure. Let D be the diffracting edge. Then AB might be the path of a corpuscle which is inflected. Further, if a corpuscle pass at a somewhat greater distance from the edge, yet also quite close, it would supposedly be repelled. The path CF represents the path of such a corpuscle. This repulsive action was supposed to diminish as the distance of passage from the diffracting edge increases, therefore a corpuscle passing farther from the edge might take the path GH. Nothing specific is said at this juncture (*Opticks*, Book **3**, observation 1) as to how the fringes might arise, but Newton expresses himself upon this subject in his famous query 3, which reads: "Are not the Rays of Light in passing by the edges and sides of Bodies, bent several times backwards and forwards, with a motion like that of an Eel? And do not the three fringes of colour'd Light above-mention'd arise from three such bendings?"

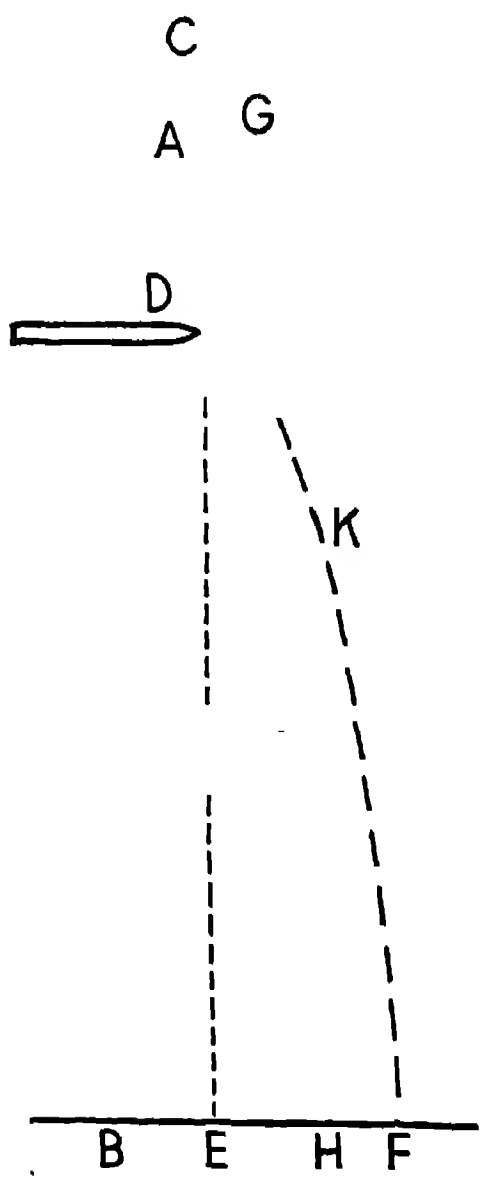


FIG. 5

Newton supposed diffraction to be due to an effect of the material of the edge; but we know today, for reasons previously pointed out, that this supposition is untenable.

A century after Newton, Fresnel urged two objections, both different from the one just mentioned, to the explanation of diffraction based upon the corpuscular theory: If we follow the locus of a given fringe, in space, from the diffracting edge backward, this locus should on the corpuscular theory be a straight line. But actually the locus of a given fringe is curvilinear. It has the form indicated by the broken curve KF. Some of Newton's

own measurements reveal this fact, and it has now been established many times. The locus is the portion of a hyperbola having the boundary of the geometrical shadow, DE, as axis and having the edge D as one of its foci. Again, on the corpuscular theory we should expect the distance of a given fringe from the edge of the geometrical shadow to remain unchanged when the source of light is brought nearer to, or removed farther from, the diffracting edge, whereas actually the distance of a given fringe from the edge of the geometrical shadow varies as markedly when the distance of the source is varied as it does when the distance of the plane of observation is varied.

Newton correctly explained the color effects which are seen in the fringes when white light is used. To obtain a clue to the origin of these colors he examined the shadow cast by a hair in light of each of the various colors of a spectrum formed by a prism. He found that as he moved the hair through the spectrum from blue to red, the external fringes moved farther from the shadow and became more widely separated. He then correctly concluded that the reason why, when white light is used, each fringe is blue on the inside and red on the outside is simply because with white light we simultaneously obtain superposed patterns of a different scale for each color—a bigger pattern for red light than for blue light.

Owing to the great weight of Newton's authority, the view that light is corpuscular came to be regarded as definitely established, and was long so regarded with a positiveness unjustified by the existing evidence. The century 1700–1800 witnessed several advances in optics—for example, the advent of the achromatic telescope—but there were only a few important advances and they were not of a type to yield new evidence regarding the basic nature of light. During this century such explanations as Newton had given of various optical phenomena, including diffraction, were retained. Even later, after a new era had dawned, many eminent scientists held to Newton's views with great tenacity.

3. The Views of Young.—The first serious attempt to revive the wave theory of light was made by the London physician, Dr. Thomas Young, who was a man of great genius. He was active in a number of fields of learning.*

During the years 1801–4 Young conducted various studies in optics. These years are now referred to as his first optical period.

* Thomas Young was born in Milverton, Somersetshire, in the year 1773. From early childhood he gave evidence of an outstanding intellect and at the age of twenty, while a medical student, wrote a paper on the mechanism of the eye which attracted attention and gained him election to the Royal Society.

After completing his medical studies and having begun to practice, Young, in 1801, became the first incumbent of the chair of natural philosophy at the Royal Institution which had been founded the year previously. (The Royal Institution was founded by Benjamin Thompson, a native of Massachusetts, who, forced to flee America at the out-

Young, from the consideration of acoustic phenomena, had conceived the principle of interference as applying to wave motion in general. He then became convinced that the so-called Newton's rings could be more rationally explained upon the wave theory of light, invoking the principle of interference, than upon the corpuscular theory. He tested his explanation in various ways, and in seeking further and independent evidence for the interference of light he was led to his now famous optical interference experiments.

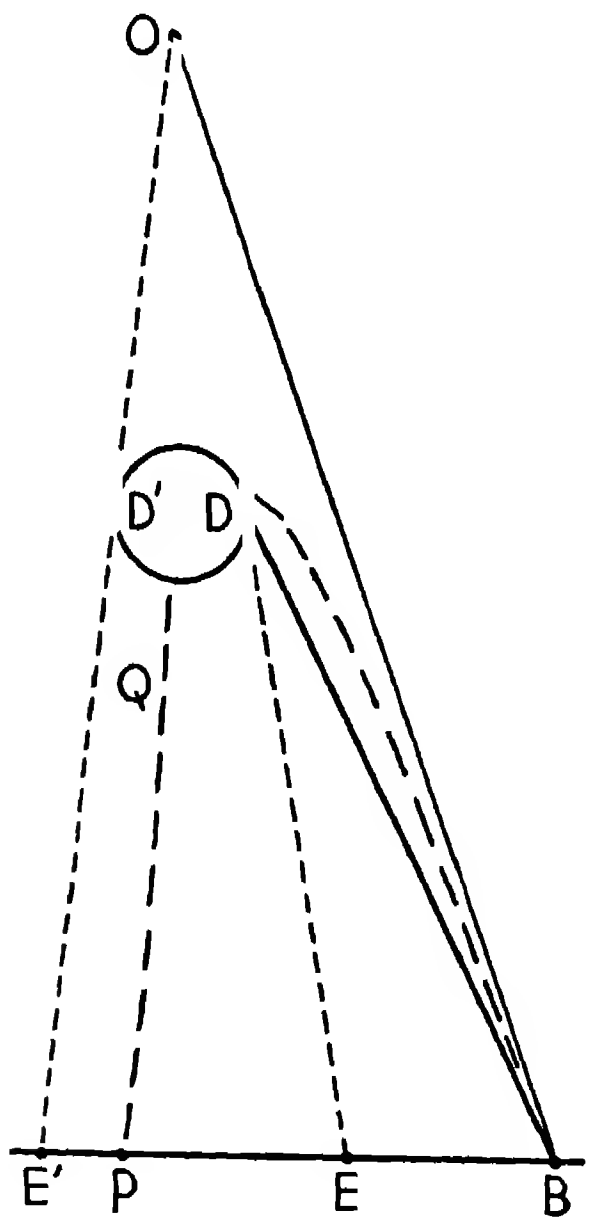


FIG. 6

Young also advanced a theory of diffraction based upon the wave theory of light and the principle of interference. This theory can best be explained by considering what one observes upon placing the eye in or near the shadow of, let us say, a narrow bar erected vertically. Referring to Figure 6, let O be the source of light, either a pinhole or a fine vertical slit, and let $D'D$ be a bar. Upon looking toward the bar from a position within the shadow such as P, one sees light coming around both sides of the bar. The very edges of the bar appear luminous over a certain length. Upon moving the eye out of the shadow the source of light comes into view, but one can still see a line of luminosity at the nearer edge of the bar and, under favorable circumstances, one can also see a faint line of luminosity at the farther edge.

The fact that light is bent into the shadow Young at first accounted for by supposing that what he called an "ethereal atmosphere" surrounds all bodies in a thin layer. A ray passing close to a diffracting edge would then be bent inward by refraction in passing through this atmosphere. He later abandoned this idea, maintaining quite correctly that the general tendency of waves to bend around obstacles sufficiently accounts for the light which is inflected. Indeed, when considering light as due to waves, the problem is rather, as Young

---

break of the Revolution because of Royalist sympathies, went abroad, and there, later, became Count Rumford.)

Young occupied the aforementioned chair for only two years and then resigned to devote himself more completely to his medical practice. He was only moderately successful as a practitioner, and he devoted the time not required by his practice to a variety of intellectual pursuits displaying great and versatile power.

He was born of good family, had a modest but sufficient private fortune, and was happily married. He attained an age of fifty-six years.

fully realized, to account for the fact that light bends around obstacles only to such a slight degree as it does.

The fringes within the shadow of a narrow bar Young supposed to be due to interference of the light coming around the two sides of the bar, and, confirming this supposition, he found that when he screened off one edge of the bar by introducing a piece of cardboard either in front of the bar or behind it, the fringes disappeared.

Observation shows, as has been stated, that the diffracted light appears to come from the edges themselves, and thus the whole situation is much as in Young's well-known interference experiment with two parallel slits. The edges of the bar, D' and D, each represent one of the two slits. As in the interference experiment, the fringes have hyperbolic loci in space, as represented, for example, by the curve PQ of Figure 6 with D' and D as foci. Young's view of the formation of the internal fringes is still acceptable today, but we must establish it on a more sound basis, discussing in particular why the inflected light should appear to come from the diffracting edges themselves—that is, why these edges should appear luminous. (This question is taken up in chap. 7, secs. 10 and 11.)

Young considered the external fringes to be formed in a manner which will be explained in referring again to Figure 6. Let the point B be the center of, let us say, a bright fringe. This fringe is supposedly formed by constructive interference of light reaching the point B by the direct path OB and light traveling from O to the edge D, being there reflected and then reaching the point B. We know now that it is not permissible to attribute diffraction to light reflected by the edge, but the fact remains that the edge appears to be luminous, it *appears* to reflect. And Young, in spite of a wrong basis for his explanation, found almost (but not quite) as good agreement with observation as the rather crude methods of measurement which he had at his disposal would be expected to yield. Incidentally it should be mentioned that in order to obtain agreement by this scheme it is necessary to assume that one-half of a period of vibration is lost at reflection (as it is in some other optical cases), and accordingly Young made this necessary assumption. Young had ascertained experimentally that the loci of the external fringes also are hyperbolic, as indicated in our figure by the curved broken line DB. The foci of the hyperbolae for the external fringes to the right of the bar are O and D, and for those to the left of the bar are O and D'. Young was able to account, upon his theory, for the hyperbolic loci in question. Further details which are pertinent in connection with Young's theory but not necessary to a general understanding are given in the footnote.*

* Offhand one would expect, according to Young's ideas, that the first bright fringe would occupy such a position that the path $OD+DB$ exceeds the direct path OB by the

Young also attempted to explain the approximate rectilinear propagation of light, but he made no real progress toward the solution of this perplexing problem.

The value of Young's various and brilliant optical investigations of the first period (1801–4) was unfortunately not appreciated for some years. His views were scathingly criticized by Lord Brougham in a series of articles in the *Edinburgh Review*. This stinging criticism, coupled with a conspicuous absence of support from other quarters, caused Young such discouragement that he decided to abandon physical science and devote himself to hieroglyphical researches. He devoted himself to these for about ten years and in this period made important contributions to the decipherment of the unknown writings of the now famous Rosetta stone.

4. Fresnel's Investigations.—Fresnel's scientific career began with suddenness and brilliance in the year 1815. In May of that year he undertook an investigation of diffraction, and on October 15 his first results were communicated to the French Academy, where they at once aroused interest. At that time Fresnel knew nothing of Young's work, and others knew little of it, although Young's papers had been published in the *Philosophical Transactions* for the years 1802 and 1804. Young's work had been ignored to such a degree that it had as yet exerted practically no influence toward reviving the wave theory of light in the world of science. The reason for Fresnel's complete ignorance of important work of a predecessor in his field of investigation will appear more fully from the sketch of Fresnel's life given in a later section. Arago, who was one of the most eminent physicists of the

---

amount λ, where λ is the wave-length of the light used. But trying this, as Young did, the calculated position falls too far from the edge of the geometrical shadow; it falls on the first dark fringe, as nearly as Young could estimate. He then assumed that one-half period is lost at reflection, thus leaving the geometrical path difference to account for only one-half period, i.e. $(OD+DB)-OB=\lambda/2$. Then, for the second bright fringe the geometrical path difference should be $3\lambda/2$, and for the third, $5\lambda/2$; etc. The positions of the fringes thus calculated agreed with observation.

That Young's scheme leads to hyperbolic loci is obvious: taking the first bright fringe as an example, we have $(OD+DB)-OB=\lambda/2$ or $OB-DB=OD-\lambda/2=$Constant, which leads to an hyperbola having the points O and D as foci.

The only measures of wave-length which Young had at his disposal were such as could be obtained from the spacing of Newton's rings or from his own interference experiments, because his investigations took place, as it may hardly be necessary to mention, before the day of the diffraction grating.

Young's views on diffraction as above outlined are given in four papers in his *Lectures on Natural Philosophy*, of which the last three are reprinted in his *Miscellaneous Works*. Passages having a bearing on his theory of diffraction occur especially as follows: *Lect. Nat. Phil.*, **1**, 458, 464, 466, 467, and **2**, 620, 629, 634, 640 ff., 645, and *Misc. Works*, **1**, 151, 164, 171, 180 ff., and 188. Much interesting material, both scientific and personal, is given in Peacock's *Life of Thomas Young*.

day, became interested in what Fresnel was doing and called his attention to the papers of Young. Subsequent study of these papers and correspondence with Young revealed that this first work of Fresnel really represented no advance in final result over what Young had done a decade earlier. However, Fresnel had shown great genius. He, having come to doubt the corpuscular theory, had arrived at the principle of interference independently in seeking an explanation for the occurrence of the internal fringes in the diffraction pattern formed by a thin rod. His explanation for these fringes was the same as that of Young, and he had likewise performed the experiment of cutting off the light which comes around one side of the rod. He at first, like Young, explained the external fringes as arising from light reflected by the diffracting edge. He too had found it necessary to assume that one-half period is lost in the process of reflection. He had discovered that the loci of the fringes in space are hyperbolic, and he had advanced this fact as an argument against Newton's corpuscular theory of light.

In addition, Fresnel had introduced an important improvement in technique. This we shall describe in referring to Figure 7. Supposing the light to come from above in the figure, let D be the diffracting object and PP' the diffusing screen upon which the pattern is received. Fresnel's first step was to replace the opaque diffusing screen which had been used by his predecessors by a piece of ground glass and view the pattern from behind the screen through a short-focus lens, L. He could thus see the pattern greatly magnified. He then found that if he took pains to place the lens at a distance from the ground glass equal to the focal length of the lens, he could remove the ground glass and still see identically the same pattern. He could see the pattern in space, so to speak. And the pattern is much brighter without the ground glass; this is in many cases a great advantage. Furthermore, it is obvious that one may just as well dispense with the diffusing screen from the outset and simply set up the lens and look through it—and this is what Fresnel finally did. One sees the identical pattern which would be formed upon a diffusing screen placed in the principal focal plane of the lens. The eye should not be placed close to the lens but should be withdrawn to approximately the rear focal point, R, in order to obtain a good field of view.

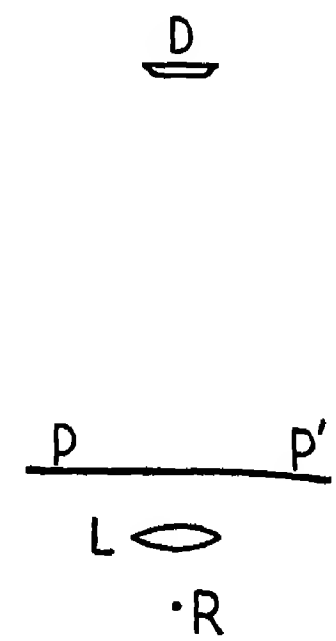


FIG. 7

The importance of this method to Fresnel was that by using an eyepiece with a crosshair in it and mounting the whole on a traveling micrometer he could make measurements vastly superior to any which could be made upon a sheet of white paper, or, following Newton, by receiving the pattern upon a white ruler. It is strange that this simple method of Fresnel's should have

waited for its introduction for a century and a half after Grimaldi's discovery of diffraction.*

Today we may of course alternatively erect a photographic plate in the plane in which we wish to obtain the pattern and allow the pattern to record itself photographically, later making measurements upon the photographic plate.

The findings of Fresnel thus far mentioned are all contained in his first memoir, which was submitted to the Academy in October, 1815. By July, 1816, he had submitted three more accounts covering further investigation. In these he reports observations which are destined to reveal to him later the insufficiency, not to say basic unsoundness, of his theory as first proposed, closely duplicating Young's theory. He discovers systematic discrepancies which are in most cases small, but he finds some quite marked ones in the pattern due to a slit, especially when this is narrow. The explanations which he devises to account for these discrepancies foreshadow a coming event, namely, the abandonment of his original theory and the replacement of this by another theory. The second theory is the one which we today mean when we speak of Fresnel's theory of diffraction. The abandoned theory is now referred to as Young's theory, because Young was the first to propound this.

Fresnel's final theory was submitted in the year 1818. In March of that year he deposited an outline of it, in a sealed envelope, with the secretary of the French Academy. The object of this procedure was to safeguard his priority while he prepared a detailed account of all of his work on diffraction for submission in a prize competition on diffraction which had recently been proposed by the Academy. The final account was submitted in July, 1818, and was awarded the prize in the following year; it is now known as Fresnel's *Mémoire couronné*.

Fresnel was the first to show that the diffraction pattern formed in a given instance is, so far as can be observed, independent of the material or nature of the diffracting edge. He demonstrated this in a series of experiments. For example, he examined the pattern due to a diffracting aperture in the form of a slit of uniform width having jaws as follows: Supposing the slit to be vertical, the left-hand jaw had along, let us say, its upper half a sharp edge and along its lower half a quite rounded edge, such as might be presented by a fairly thick bar, whereas the right-hand jaw was rounded

* The method became generally known after Fresnel's introduction of it, and Fresnel was the first to make use of it for the important purpose of enhancing the accuracy of measurement. However, the astronomer Delisle had used the method for purely observational purposes a century earlier, but its value was evidently not recognized by others and it had been completely forgotten. See Delisle's paper in *Acad. Roy. des Sci.* (1715), on p. 168, and his *Memoirs* (St. Petersburg, 1738), p. 213.

along its upper half and sharp along its lower half. If now there were an effect of the nature of the edge, one would expect the pattern formed by this slit, when using a point source of light, to be in some way different in its upper and lower halves. Instead, the two halves were found to be identical. Each fringe was uniformly bright and perfectly straight over its entire length.

References to Fresnel's writings.—The following passage in small print should be omitted by most readers. It is inserted for reference, for the benefit of anyone who may wish to acquaint himself at first hand with the work of Fresnel.

The outstanding parts of Fresnel's *Mémoire couronné* may be found in English in *The Wave Theory of Light,* edited by Crew, which is a volume of the series, "Scientific Memoirs," edited by Ames (American Book Co.). The important parts of three of Young's papers are also reprinted in the same volume.

Fresnel's writings and related material were collected and published under the auspices of the French government about forty years after his death, in three large quarto volumes entitled *Œuvres complètes d'Augustin Fresnel.* The material on diffraction comprises the first half of Volume **1.**

Volume **3** contains an abstract of the material of all three volumes, beginning on page 529. In addition, each volume has its own Table of Contents at the rear.

The Introduction by Verdet at the beginning of Volume **1** contains much of general interest. An account of Fresnel's life by Arago is given in **3,** 475. Scientific correspondence is given in **2,** 737, and this includes numerous letters exchanged with Young.

There are six principal papers on diffraction (including interference incidentally). These occur in Volume **1** as follows:

FIRST GROUP: COMPRISING FOUR PAPERS

1. *First Memoir,* page 9, presented to the Academy on Oct. 15, 1815; not published at the time except as explained under 3.

2. Supplement to the *First Memoir,* page 41, presented to the Academy on November 10, 1815; also not published at the time except as under 3.

3. *Second Memoir,* page 89; this is a revision of Nos. 1 and 2 for publication in the *Ann. de chim. et de phys.,* where it appeared in March, 1816. It repeats practically the entire contents of Nos. 1 and 2 but also contains additions, especially more accurate measurements.

4. Supplement to the *Second Memoir,* page 129, presented to the Academy on July 15, 1816; apparently not published at the time.

SECOND GROUP: COMPRISING TWO PAPERS

5. "Note on the Theory of Diffraction," page 171, this being the outline of the final theory, deposited, sealed, with the secretary of the Academy, on April 20, 1818.

6. *Mémoire couronné,* page 247, deposited at the Academy on July 29, 1818; awarded the prize in March, 1819. An extract was published in *Ann. de chim. et de*

phys., Volume **11** (July and August, 1819). Published in full for the first time in 1826 in Volume **5** of the *Mém. de l'Acad. Roy. des Sc.* for the years 1821 and 1822.

There is much material in addition to the foregoing principal papers. This consists of notes subsequently appended to the memoirs, correspondence, reports by the committees to whom the memoirs were referred, etc.

The prize memoir, 6, repeats much of the contents of all of the previous papers, and since 3 is already a repetition of 1 and 2, there is much repetition. The total bulk of the material is great and it is consequently difficult to find any one thing sought for—even with the help of the abstracts given in Volume **3**. In view of this fact, it may be worth while to note where some items of especial interest occur; we shall give in each case what is presumably first occurrence. In most cases there is subsequent repetition. The references are to Volume **1**.

On page 13, viewing the pattern on ground glass, through a lens, and then with a lens alone; page 14, a method suggested to Fresnel by Arago of forming a point source of sunlight by the aid of a short-focus lens. Fresnel varied this by using a drop of honey (the object of the method is to facilitate using sunlight without a heliostat); page 16, disappearance of internal fringes upon cutting off light coming around one edge of a rod; page 17, principle of interference, a very imperfect first conception; page 19, hyperbolic loci of external fringes; page 26, half-period loss of phase at reflection; page 28, lack of influence of nature of edge (again on p. 173); pages 45 ff., foreshadowings of the diffraction grating; page 150, description of Fresnel mirrors; page 160, the first definite foreshadowings of Fresnel's final theory of diffraction.

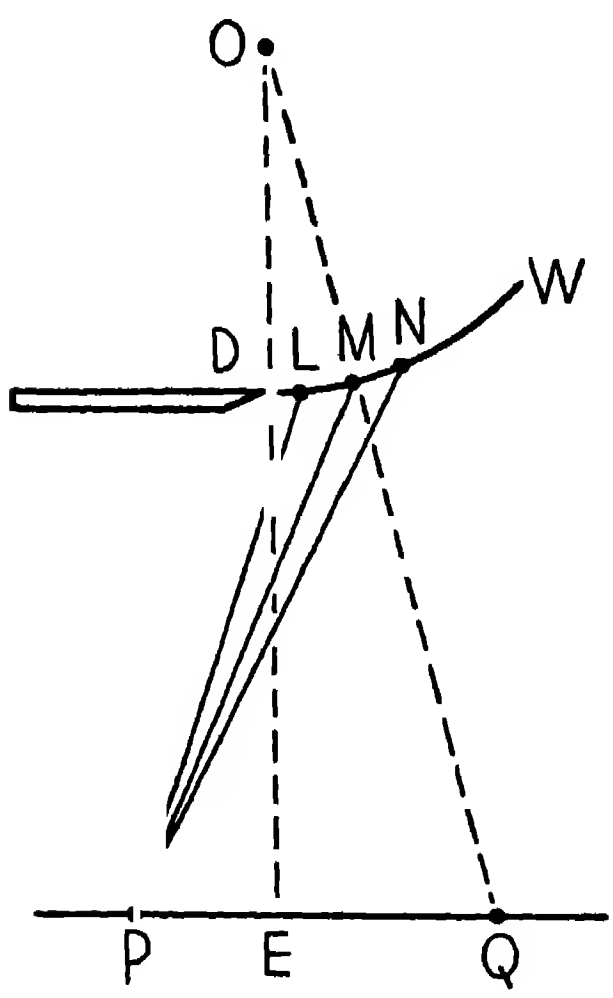


FIG. 8

On page 173 the lack of influence of the nature of the diffracting edge is for the first time systematically established (that Young considered the nature of the edge to have an influence is evident when he says, "the fringes are however rendered more obvious as the quantity of this reflected light is greater" [*Lect. Nat. Phil.*, **1**, 467]); page 174, introduction of Huygen's principle; page 176, Fresnel integrals; page 302, anticipation of Fraunhofer diffraction; page 330, Fresnel biprism; page 365, Fresnel zones.

5. The Principles of Fresnel's Theory.—According to Fresnel's final theory, the diffracting system is to be regarded simply as a barrier which obstructs a portion of the wave and allows the remainder to pass. Referring to Figure 8, let O be a point source of light, let D be a diffracting edge, and W a wave front which is just passing the edge. Suppose the wave is one of a long, simple harmonic train; in this case a simple harmonic oscillation is taking place at each point of the wave front, for example, at the points L, M, and N. The oscillations are all in the same phase. We must now invoke

Huygens' principle. That is we must consider each point of the wave front to be the source of an elementary wave, a so-called "wavelet." Let P be a given field point in the plane PQ in which we wish to deduce the distribution of illumination. The disturbances all originate upon the wave front in the same phase but have different distances to traverse to reach P and we must sum up the disturbances reaching P in accordance with the principle of interference. The disturbance reaching P from L will be ahead, in phase, of the disturbance from M, which will in turn be ahead of the disturbance from N. The resultant effect at P is to be found by summing up the contributions from all points of the effective portion of the wave front, taking due account of the phase in which each disturbance arrives.

The distances from the point P to the points L, M, and N differ markedly from each other, and hence the disturbances reaching P from L, M, and N will differ markedly from each other in phase. However, the distances from some point such as Q to L, M, and N differ only slightly from each other, and hence the disturbances from L, M, and N will reach Q differing only slightly in phase from each other. It is thus obvious that the resultant effect found for Q might be much greater than that found for P—and so it will turn out to be when we learn to make the summation properly—for the whole effective portion of the wave front, as we shall learn in later chapters.

Fresnel's theory accounts for the distribution of illumination which actually obtains in any given plane PQ. It thus accounts for the observed diffraction pattern and, at the same stroke, accounts for the approximate rectilinear propagation of light. The theory, when worked out, shows that the amount to which waves bend around obstacles becomes less and less as we pass to waves of smaller and smaller wave-length. It is the extreme shortness of the wave-length which turns out to be the reason for the approximate rectilinear propagation of light.

When both external and internal fringes occur, as in the pattern for a thin rod, we no longer need to seek separate explanations for the two kinds of fringes. The theory, applied according to the principles outlined, accounts, in one operation, for both kinds of fringes.

Although Fresnel introduced Huygens' principle as a part of his theory of diffraction, he did not adopt all of Huygens' views regarding the nature of light. Huygens believed light to be due to single random pulses propagated through the ether, whereas Fresnel explicitly stated his belief that light involves long trains of waves—and this, with certain qualifications, is what we still believe today (the qualifications are discussed in chap. 7, secs. 22, 23, and 24).

Huygens' attempt to explain rectilinear propagation also requires comment. He tried to explain *strict* rectilinear propagation. He was unaware, at least in 1678 when he wrote his *Traité de la lumière*, of the deviations

from rectilinear propagation which actually occur—he was unaware of diffraction. To explain strict rectilinear propagation Huygens advanced an argument purporting to prove that each elementary wavelet will be effective only at the point which lies upon the common envelope of all of the elementary wavelets. This argument begs the whole question and does not in any way account for the deviations from rectilinear propagation which actually occur. The argument is unsound and must be simply ignored.

Fresnel's theory of diffraction applies equally to longitudinal waves or to transverse waves. Huygens and all of those who in the early days believed in the wave theory of light took for granted that light-waves are longitudinal, and this belief carried over into the days of Young and Fresnel. The idea that light-waves are transverse, which has by now been so abundantly confirmed, was propounded by Young and by Fresnel independently in order to explain polarization. This hypothesis constituted an important advance in physics. More will be said later about the circumstances of its introduction. Suffice it for the present to state that when Fresnel wrote his final memoir on diffraction, in the first half of the year 1818, he wrote it still supposing that light-waves are longitudinal. But this fact does not invalidate the theory which he there developed, because the theory itself applies equally well to transverse waves.

By way of recapitulation, the basic principles of Fresnel's theory of diffraction are: (1) that the diffracting object acts simply as a barrier; (2) that light involves long trains of waves; (3) that we must apply Huygens' principle at each point of the wave front; and (4) that we must apply the principle of interference in summing up the disturbances which reach a given field point. Fresnel was the first to apply *Huygens' principle and the principle of interference in conjunction with each other.*

6. Sketch of the Life of Fresnel.—Augustin Jean Fresnel, the son of an architect, was born at Broglie, near Caen, in Normandy on May 10, 1788. He was somewhat backward intellectually during his early years and was frail in physique during his entire life. After preliminary schooling at Caen he entered the Ecole Polytechnique in Paris in 1804 and then continued in the Ecole des Ponts et Chaussées. Upon graduation he entered government service as an engineer and was sent to outlying districts where his duties consisted in supervising work on roads and irrigation ditches, the occasional building of small bridges, etc. This work was little to his liking, but he remained at his task. During leisure hours he speculated upon philosophical and scientific questions and never gave up hope of some day engaging in scientific activity. He procured such books and current scientific literature as he could obtain, and it is known that early in 1814 while stationed at Nyons in the south of France he was speculating anew in regard to difficulties in the then prevalent doctrines of the material nature of light and heat.

In March, 1815, when Napoleon returned from Elba, Fresnel, who favored the royalist régime, joined a small force under the Duc d'Angoulème which vainly offered momentary resistance to the advance of Napoleon on his way to Paris to reassume power. Soon thereafter Fresnel was suspended from his governmental appointment as engineer, because he had joined the force which had attempted to resist Napoleon, and he was placed under police surveillance. He was allowed, however, to join his mother who, after her husband's death, had moved to Mathieu, also near Caen in Normandy. During his enforced leisure at Mathieu, Fresnel turned to experiments on diffraction, hoping that these might lead him to a more satisfactory conception of the nature of light. He began his work without scientific training, except such as he had received at the engineering school, and after having been away from centers of learning for six or eight years.

After the Battle of Waterloo, June 18, 1815, and the shortly following second restoration of Louis XVIII, Fresnel was reappointed as government engineer and an assignment was arranged which allowed him to remain in or near Paris; but his scientific work had to be carried on entirely outside of his official duties. He at one time sought to exchange his post as engineer for a vacant post as examiner at the Ecole de la Marine because this would have allowed him more time for scientific pursuit. The selection of Fresnel for the vacant post seemed assured when, at the last moment, political expediency led to the appointment of someone else.

In June, 1819, on the recommendation of Arago, Fresnel was appointed to serve with the Commission on Lighthouses and he soon revolutionized the design of coastal beacons. He replaced the reflectors which were then in vogue by cellular lenses. His designs for these are still in use. In 1824 Fresnel's health began to fail seriously. He was forced to restrict his hours of labor and, feeling in duty bound to devote what strength he could muster to his work on the improvement of lighthouses, he thenceforth abandoned scientific investigation. He had contracted tuberculosis, and he died three years later, on July 14, 1827, having attained an age of only thirty-nine years. His creative scientific life was limited to nine years, 1815–24, but his achievements during this short period were tremendous. His major contributions were all in the field of optics. He founded the theory of diffraction and carried it far. He made epoch-making advances in the theory of polarization and double refraction; these are considered by many as his greatest work. He also made important advances in the theory of reflection and of ordinary refraction, and in the optics of moving mediums.

7. Development of the Wave Theory of Light.—The process of reviving the wave theory of light in the scientific world was a long and involved one. Let us next consider the part which the phenomena of double refraction and polarization played in this process.

The double refraction of Iceland spar, now commonly known as calcite, had been discovered by Erasmus Bartholinus of Copenhagen in 1669. This phenomenon is illustrated in the upper part of Figure 9, where, on the left, i represents a ray of light incident upon a rhomb of calcite. This ray is divided into two rays of which one, the ordinary ray, O, is refracted in the same manner as a ray would be refracted in entering, for example, a block of glass, and the other, the extraordinary ray, E, is refracted as indicated.

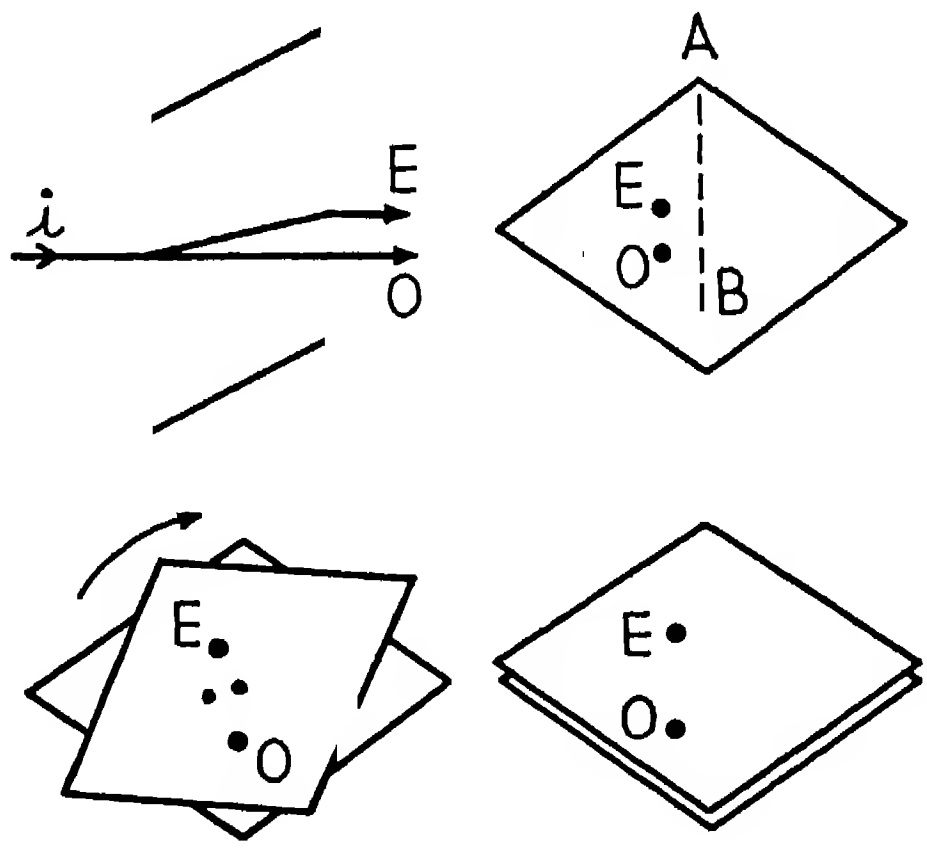


FIG. 9

Upon looking through the crystal at, let us say, an ink dot upon a piece of paper, one sees two images of the ink dot as illustrated in the upper right-hand corner of the figure by the dots O and E. The two images are seen with equal clearness. They lie along a line parallel to the line AB which by construction bisects the obtuse angle formed by the edges of the crystal face at A; the line AB, so constructed, forms the short diagonal of the crystal face when this face is equilateral as illustrated and as we shall henceforth suppose it to be.

Huygens, in his *Traité de la lumière* (1678), had described a further discovery in the same connection. This is illustrated in the lower part of our figure: When two crystals of calcite are placed on top of each other, as shown on the left, then each of the two images formed by the first crystal is again divided by the second crystal into two images lying along a parallel to the short diagonal of the second crystal. The ray, which is refracted in the ordinary manner in both crystals, gives rise to the image O, and the ray which is extraordinarily refracted in both crystals gives rise to E. Of the two remaining images the one on the left is due to the ray which is ordinary in the first crystal and extraordinary in the second, and vice versa for the image on the right. The images O and E have equal intensity, but in gen-

eral a different intensity from the other two images, which also have equal intensity. In the relative positions of the two crystals illustrated the images O and E are the more intense ones. If now the topmost crystal is rotated in the direction indicated by the arrow, the fainter images become still fainter and disappear when the two crystals attain the relative positions shown at the right of the figure, with corresponding edges parallel. The two beams into which the light is divided in passing through the first crystal have each evidently acquired some new property which causes the second crystal to divide each beam into two beams having relative intensities which depend upon the orientation of the second crystal with reference to the first.

The new property, whatever it may be, is called "polarization." Huygens was at a loss to explain polarization on his longitudinal wave theory. Today we consider light as due to transverse waves—that is, we believe that the vibration takes place entirely in the plane perpendicular to the ray, and we believe that in ordinary light, in the light before passing through the first crystal, the vibration changes its direction in the plane perpendicular to the ray many times per second. Further, the polarizing action of the first crystal supposedly consists in resolving the original vibration, frequently changing in direction, at every stage into two components. One component is refracted as the ordinary ray; this component is parallel to the long diagonal of the crystal face. The other component is refracted as the extraordinary ray; this component lies in the plane containing the short diagonal of the crystal face. In chapter 7, section 17, the evidence that this is what occurs will be presented. This evidence is comparatively recent. However, long before there was any way of deciding in which of the two planes the vibration for each ray takes place, convention led to calling the other plane the "plane of polarization" for that ray. Consequently, according to this terminology of long standing which is still used today, the plane of polarization is the plane which is perpendicular to the plane in which the vibration takes place, or perpendicular to what is now called the "vibration plane."

The second of the two calcite crystals in its turn resolves each of the two vibrations which leave the first crystal into one component parallel to the long diagonal of the second crystal and into another which lies in the plane containing the short diagonal.

Newton in commenting on Huygens' discovery of polarization said that it shows that "the Rays of Light have different Properties in their different Sides," and also, "it's difficult to conceive how the Rays of Light, unless they be Bodies, can have a permanent Virtue in two of their Sides which is not in their other Sides . . . ." (*Opticks*, Book **3**, queries 28 and 29). Since the phenomenon of polarization had proved to be totally inexplicable on Huygens' longitudinal wave theory, the foregoing fragmentary explanation of Newton was seized upon, and the phenomenon of polarization, as well as

that of rectilinear propagation, came to be regarded as favoring the corpuscular theory.

About a century after Huygens and Newton it was discovered that light could be polarized in another way than by passing it through a crystal of Iceland spar. The French military engineer Malus discovered in the year 1808 that light can be polarized by reflecting it at the proper angle from glass or from water. This discovery removed the property of polarization from the realm of the anomalous (having previously been producible only by passing light through a crystal) to the status of a more important and general property of light, and, since polarization was still wholly inexplicable upon the (longitudinal) wave theory, Malus' discovery was regarded as furnishing new strong evidence for the corpuscular theory. A shortly following experimental investigation by Biot and also a mathematical paper by La Place were regarded as still further strengthening this evidence. The years after Malus' discovery and a decade or thereabouts after Young's work have been described as the darkest ones in the history of the wave theory.*

The first work of Fresnel on diffraction, in 1815, aroused the interest of Arago and he then sought out Fresnel in order to make his acquaintance. The two men became and remained close friends. Arago had previously been interested in Young's work although he was not thoroughly acquainted with it. He too now came definitely to favor the wave theory, and thereupon the rivalry between the two theories became active in France. Young's work was then for a time much better known in France than in England. The corpuscularists, however, still held the upper hand. They were led by La Place, Poisson, and Biot. Discussion at times became bitter. It led to a quarrel between Arago and Biot who had been close friends, and their friendship was never resumed.

In England, Brewster had been making important contributions to optics—and he favored the corpuscular theory. In the meantime Young had resumed activity in the field of optics, this activity taking the form of writing critical reviews and summaries of the optical developments of the day.

Toward the close of the year 1816 both Young and Fresnel independently had the first inkling that the perplexing phenomenon of polarization might yield to explanation, upon the wave theory, if by any chance light-waves might be transverse, but the very idea of transverse light-waves seemed preposterous from the point of view of the day. Four or five years elapsed. Then the hypothesis that light is due to purely transverse waves was defi-

* Whewell, *History of Inductive Sciences* (New York, 1858), **2**, 100.

nitely advanced by Fresnel. We cannot appreciate in retrospect how difficult the new conception was.*

8. Elastic Solid Theory of the Ether.—Waves in the ether (thought of as an elastic medium) had always meant one thing and one thing only, namely, longitudinal waves. The theory of elasticity had not been developed far, and wave motion was not any too well understood. The ether had been thought of, often without full realization, as a fluid, and it was taken for granted, correctly enough while thinking of the ether as a fluid, that the waves would be longitudinal.

A fluid has volume elasticity only, and it can hence propagate only compressional waves, which are longitudinal. A solid, on the other hand, has both volume elasticity and elasticity of shape; it propagates compressional—that is, longitudinal waves—and also distortional—that is, transverse waves.

The conception of transverse light-waves involved in the first place getting away from the conception of the ether as a fluid and conceiving it as a solid. The difficulty of this was in itself very great. And added to this was the difficulty of accounting for the longitudinal waves which a solid ordinarily propagates along with transverse waves. In light, these longitudinal waves appear to be absent because otherwise we could not polarize a beam, for example by a Nicol prism, and then extinguish it completely by crossing a second Nicol prism with the first one.

It was in the year 1821 that Fresnel propounded the hypothesis that light is due to waves which are exclusively transverse, pointing out how the extensive observations which he and others had made could be harmonized by this hypothesis. From then on the balance of evidence began to turn rapidly in favor of the wave theory. The ether soon came to be regarded definitely as an elastic solid, but there remained the question of what becomes of the longitudinal waves. This question occupied many of the eminent minds of the day—and of even much later days. Lord Kelvin treated it extensively as late as the year 1888.

The year 1821 is notable for us for another reason. It is the year in which the young German optician Joseph Fraunhofer reported his observations upon a class of diffraction phenomena which have now been named after him, and, further, reported his discovery of the diffraction grating. He had already several years previously reported the discovery of the dark lines in the solar spectrum which we now call the "Fraunhofer lines."†

* Pertinent historical references are: Young, *Misc. Works*, **1**, 332 and 383; Fresnel, *Œuvres*, **1**, 294, 394, 527, 629, 634, and 635; Whewell, *op. cit.*, **2**, 100 ff. and 106; Whittaker, *History of the Theories of Aether and Electricity*, pp. 121 and 123 ff.; Cajori, *History of Physics*, p. 154.

† Fraunhofer made each of the advances mentioned entirely independently—and brought each to fruition. However, each advance had been to some extent previously re-

It is to Fraunhofer more than to any one other man that we are indebted for introducing the spectroscope, and, further, his discovery of the grating made possible for the first time the measurement of wave-lengths in the spectrum with any real accuracy. His work also added further evidence in favor of the wave theory.

For many years following, however, several eminent scientific men believed that a final decision between the corpuscular and wave theories could be obtained only when it would be found possible to perform a certain *experimentum crucis*. The experiment which they had in mind was attended, however, by experimental difficulties which had thus far proved insurmountable:

According to the wave theory, a ray of light entering, let us say water, is bent toward the normal to the surface because the waves travel more *slowly* in the refracting medium. The proof of this by Huygens' construction is today familiar to all. According to the corpuscular theory on the other hand, light should travel *faster* in a refracting medium than in air of vacuum. This is deduced on the ground that a corpuscle just before entering a refracting medium would be attracted by the medium and thus caused to travel faster within the medium. This attraction would affect only that component of the original velocity which is normal to the surface and the attraction would thus account for the bending-in of the ray toward the normal—in accordance with the observed fact. Accordingly, the crucial experiment which was awaited was a direct determination of whether the velocity of light is greater or less in a refracting medium than in air or vacuum. The velocity of light had as yet been determined only by Römers' and by Bradley's astronomical methods. In 1849 the first direct determination was made, by Fizeau, but the method necessitated a working distance of several kilometers. Then in 1850 Foucault devised a method which required a path of only some twenty meters. By placing a tube of water in the path he was able to show that the velocity of light is less in water than in air. The determination of this fact by Foucault rendered a decision in favor of the wave theory, and the corpuscular theory was thereafter abandoned—until 1900, when it was revied in a new form. We shall explain later why it was revived.

9. Advent of the Electromagnetic Theory.—The electromagnetic theory of light is also a wave theory. Its innovation consists in establishing a relation between light, and electricity and magnetism.

---

ported, but, not being followed up, had borne no fruit. The presence of a few dark lines in the solar spectrum was observed by Wollaston in 1802 (for details see, e.g., Cajori, *op. cit.*, p. 162); the Fraunhofer class of diffraction phenomena were clearly anticipated by Fresnel (see *op. cit.*, **1**, 302; a diffraction grating had been made and described by Rittenhouse, *op. cit.*, **2**, 201, 1786; attention was recently directed to the Rittenhouse grating by Cope, *Jour. Frank. Inst.*, **214**, 99, 1932.

In the year 1820 Oersted made the now famous discovery that a wire carrying an electric current exerts a force on a magnetic needle, thus revealing for the first time that a connection exists between electricity on the one hand and magnetism on the other. Then followed Ampere's series of researches in which he showed that a coil of wire carrying a current behaves in every way like a magnet, and in 1831 Faraday and Joseph Henry, independently, showed that a magnet induces a current in a closed circuit while the magnet is being brought up to the circuit or while it is being withdrawn from it.

Not many years after these great discoveries, which revealed the connections which exist between electricity and magnetism, Faraday began to speculate upon whether there might not be a connection also between electricity and magnetism, considered jointly, on the one hand, and light on the other. After many fruitless experiments to establish such a relation Faraday finally succeeded. He placed a piece of glass in a magnetic field formed between the poles of a large electromagnet and then passed a beam of polarized light parallel to the lines of magnetic force. He found that the magnetic field caused the plane of polarization of the light to rotate in advancing through the glass. This effect has since come to be known as the "Faraday effect," and it long stood as the only known connection between light, and electricity and magnetism.

During the years 1865–70 Clerk Maxwell developed electromagnetic theory to a point which led him to the supposition that there might be such a thing as "electromagnetic waves," whatever these might be. But nobody knew how to go at trying to produce these waves. However, the velocity of these mysterious electromagnetic waves could be calculated, and the calculated velocity turned out to be equal to the velocity of light, which had been experimentally determined. Theory also indicated that the electromagnetic waves should be purely transverse, and this fact suggested that light might be some form of electromagnetic wave. The theoretical researches of Maxwell thus furnished new and strong evidence of an intimate relationship between light and electricity and magnetism, but the situation remained rather obscure for nearly twenty years longer.

In 1888 Heinrich Hertz began to experiment with a view to producing Maxwell's electromagnetic waves in the laboratory. He succedeed in producing the waves by a so-called "oscillator." This may be represented, as in Figure 10, by two vertical rods *RR* placed end to end and separated by gap as illustrated. One rod—let us say the upper one—is charged positively, and the other negatively, by an induction coil *C*. As the difference of potential between the upper rod and the lower is increased, the value of the electric intensity in the space surrounding the rods rises. When this intensity attains a sufficiently high value, the insulating power of the air breaks

down and a spark passes between the rods. During the passage of the spark the charge on each rod passes to the other rod, then back again, then to the other again, etc. That is, there is a surging or oscillation of the charge. The charge, in passing to and fro from one rod to the other, constitutes an oscillating electric current which in turn gives rise to an oscillating magnetic field at right angles to the rods forming the oscillator. There are electro-

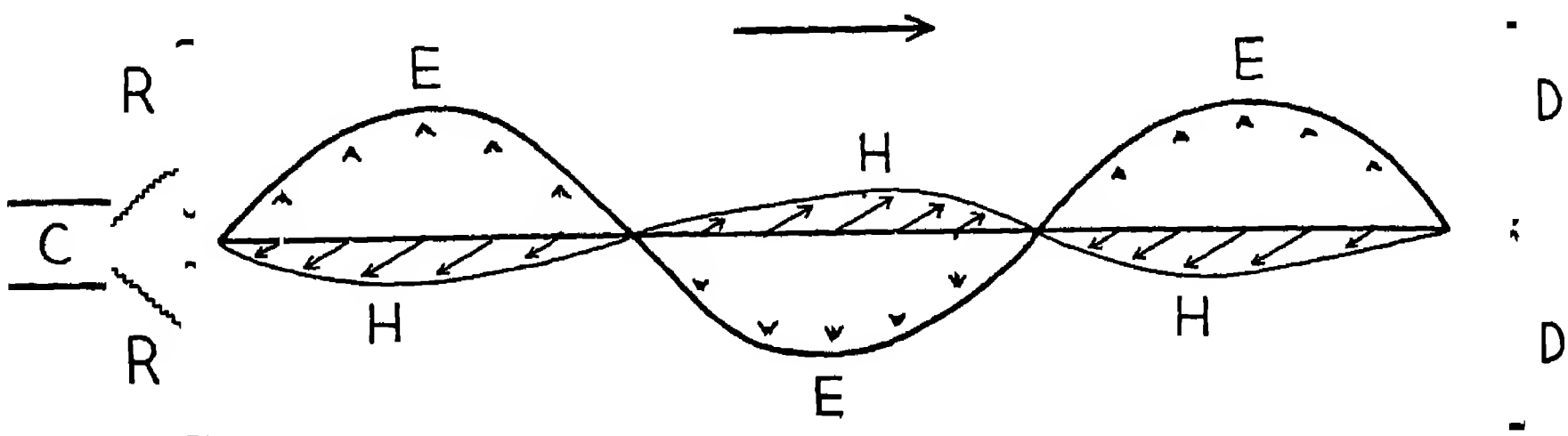


FIG. 10

magnetic waves sent out. For a direction of propagation horizontally to the right, indicated by the arrow above the figure, the waves take the form illustrated. The undulating electric force, E, lies in the plane of the figure and is vertical, and the undulating magnetic force, H, is perpendicular to the plane of the figure; it is horizontal. Both vectors are perpendicular to the direction of propagation of the waves—that is, the waves are transverse.*

Hertz detected the existence of the waves by a second oscillator, which we shall again represent as two rods, DD. The gap between the two rods is very small in the detector. The incoming waves build up sufficient potential difference between the two portions of the detector to cause a minute spark to pass in the small gap. This spark reveals the existence of the waves—which induce a surge of charge in the detector.

Having devised the above-described means for producing and detecting electromagnetic waves, Hertz made large mirrors of metal and showed that the waves could be reflected—and he brought the waves to a focus by a concave mirror. He then made a large prism of pitch and showed that the waves could be refracted. He also made large lenses of pitch. In short, he reproduced with electromagnetic waves many of the common experiments of optics.

Maxwell's theory and Hertz's experiments gave rise to the following conception of the origin and nature of light. A minute electric charge, which we today identify with the electron, executes periodic motion within an

* For the sake of simplicity our figure represents the waves as constant in amplitude, whereas under the given conditions the amplitude would vary along the train. This feature of the representation should not be taken literally.

atom. Referring to Figure 11, let e represent the electron, and let n represent the vastly more massive nucleus of the atom. The electron is in some manner bound to the nucleus. Let us now suppose that by some agency such as a collision with another atom, the electron is set into vibration and executes oscillations along the short vertical line indicated. This is a highly

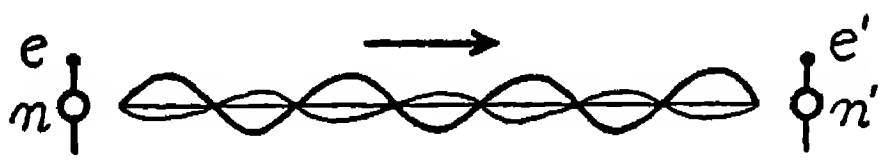


FIG. 11

special type of vibration but may serve to illustrate. The electron so vibrating constitutes what is called a "linear oscillator" and sends out waves resembling those in the previous figure. The oscillator is, however, now of very minute dimension; the oscillations are extremely rapid and consequently the wave-length is extremely short. The waves which are sent out are light-waves, and when these strike the retina of the eye or a photographic plate they set into vibration another electron, e', bound in another atom, n'—and this vibration in some manner produces the sensation of light, or the darkening of the photographic plate.

It is hoped that the above-painted scant and crude picture may convey to the uninitiated reader some idea of what is meant by Maxwell's "electromagnetic theory of light."

The success of the electromagnetic theory of light gave further grounds for the firm belief that light is due to wave motion. So firm was this belief that during the closing decade of the last century several eminent physicists remarked that if we know anything certainly, we know that light is due to waves and *is not* due to corpuscles. But this feeling of certainty was destined to be short lived.

10. Recent General Developments.—In 1900 Max Planck, in seeking to explain the distribution of the intensity of radiation from an incandescent black body at various wave-lengths, was forced to the hypothesis that light must be emitted in "quanta." Subsequent extension of this hypothesis to other phenomena—for example, to the photo-electric effect—led to the conclusion that *absorption* must also occur in quanta. When it became evident that both emission and absorption take place in quanta, it was at first found difficult to reconcile this circumstance with the wave theory of light. Absorption especially presented difficulty. But both emission and absorption in quanta are self-evident upon the corpuscular theory of light, and hence the corpuscular theory was revived to explain quantum phenomena. For some years it appeared as though the two theories, which were old rivals, and which it was now necessary to believe in simultaneously, were in antithesis to each other, but today it has been found that the two apparently con-

tradictory views can be reconciled. It, however, continues to be a convenience to explain some groups of phenomena upon the wave theory and to explain other groups upon the corpuscular theory.

The phenomena of diffraction remain among those phenomena which are to be interpreted on the wave theory.

The advent of the electromagnetic theory of light removed the necessity previously felt by many of forming a purely mechanical conception of the ether, but the ether itself retained its full reality in this theory into the present century. However, the theory of relativity, propounded by Einstein in the year 1905, shook the previously firm belief in the existence of the ether, and this belief remains shaken today.

Fortunately many of the phenomena of optics, including those of diffraction, can be interpreted without making a decision one way or the other about the existence of the ether. The phenomena in question may be explained on the basis of wave motion, and we do not need to answer the questions "Waves of what?" and "Waves in what?" We may speak simply in terms of the "disturbance" which occurs at a given point in space without philosophizing as to the ultimate nature of this disturbance.

11. Developments in Diffraction Theory.—From the time of Young, Fresnel, and Fraunhofer to the present day, extensive development has occurred. Let us for present convenience regard any portion of diffraction theory as falling into one of two categories:

Let the first category comprise those portions of the subject which involve applications and extensions of the basic principles propounded by Fresnel in his final theory of diffraction and little or nothing further as regards fundamental concept. A great majority of diffraction phenomena fall into this category. This category includes the so-called Fresnel class of diffraction phenomena, the so-called Fraunhofer class of diffraction phenomena, the important "principle of Babinet," the applications of diffraction theory to the theory of optical instruments, in particular the extensive theory of the grating, etc. The discussion of the various matters of this category perforce constitutes a predominating portion of an introductory treatment of diffraction such as this book aims to present.

As to the second category, namely, the developments involving theory which basically transcends the principles laid down by Fresnel: These developments consist for the most part of elaborate mathematical treatments and constitute the so-called "dynamical theory" of diffraction. The purpose of these treatments is to erect a structure which is more rigorous than Fresnel's theory. Fresnel's theory is often referred to by way of distinction as the kinematical theory of diffraction. The dynamical treatments are much more satisfactory from a general or philosophical point of view. However, they add only moderately to what Fresnel's theory accomplishes much

more simply by way of providing satisfactory explanation of optical phenomena of interest and importance. The developments of this, our second category, will also be considered in the present book, but only rather briefly. The discussion of these developments will be prefaced by showing wherein Fresnel's theory lacks rigor. The closing sections of chapter 7 are devoted to the various matters which here come into question, beginning with section 25.

There have also been important developments purely of method. For example, the graphical method of treating diffraction was first introduced by A. Cornu shortly after the year 1870. This is a method of great power and we shall employ it wherever feasible.

CHAPTER 2

THE METHOD OF ZONES

The method of zones offers the easiest approach to diffraction theory. Historically, however, this method was introduced as an afterthought. It was introduced after the more difficult method which forms the subject of chapter 3 had been completely developed.

1. Fresnel Zones.—Let the point O, in Figure 12, represent a point source of light which emits a simple harmonic train of waves of wave-length λ. Such a train constitutes "monochromatic light," which is well approximated, for example, by the light from a mercury arc when a color filter is employed which passes only the green, because the green light of the mercury arc all originates from a single spectrum line. Any sphere taken about O may represent the instantaneous position of a wave front. Accordingly, let the arc M_0M represent the intersection of a wave front with the plane of the drawing. Let P be an arbitrarily chosen point at which we wish to determine the illumination under given conditions, a so-called "field point." The line OP intersects the wave at the point M_0, which is known as the pole of the wave front with reference to the field point P. Denote the distance OM_0 by a and the distance M_0P by b. Suppose that M_1 is a point on the wave front such that the distance $M_1P \equiv r_1 = b + \lambda/2$. Rotate the line M_1P about OP as axis. The point M_1 then describes a circle on the wave front. The area of the wave front lying within this circle constitutes the first "Fresnel zone." The second zone extends from the outer edge of the first zone to points on the wave front lying at a distance $M_2P \equiv r_2 = b + \lambda$ from P. The third zone extends from the outer edge of the second zone to points lying at a distance $M_3P \equiv r_3 = b + 3\lambda/2$ from P. The outer edge of each zone lies at a distance $\lambda/2$ farther from P than does its inner edge, and upon this basis the construction is to be carried to zones of higher number.

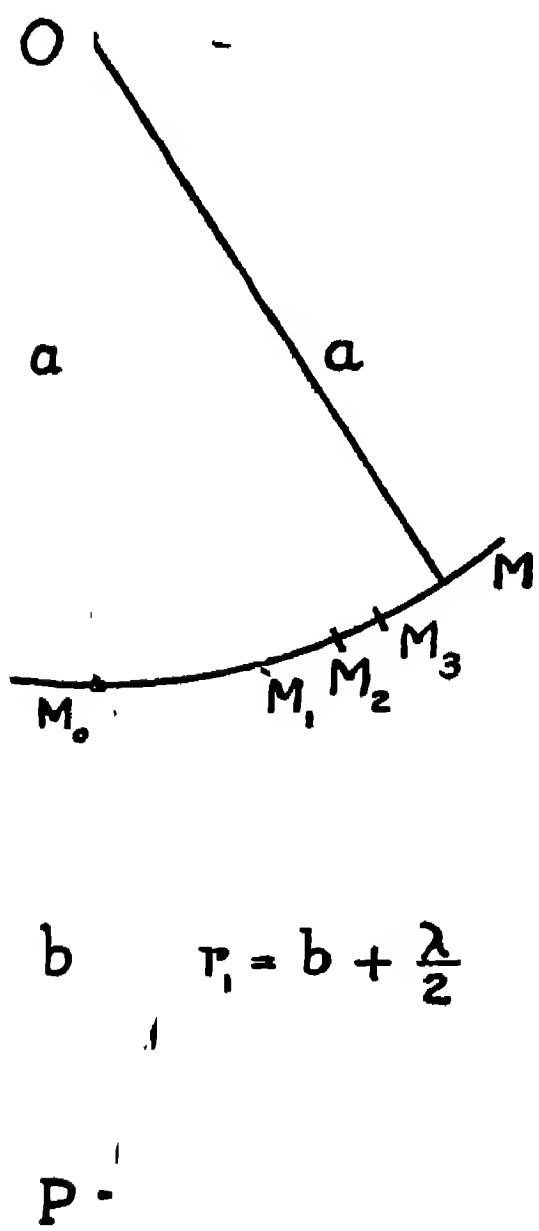


FIG. 12

This scheme of considering the wave front divided off into zones was

originated by Fresnel,* as the name given to the zones implies. The usefulness of the scheme will become apparent later. Fresnel zones were formerly called Huygens zones, but this was improper nomenclature which has now been corrected.

2. Retardation, or Path Increment.—We shall have need for an expression which relates the length of the semichord ρ (Fig. 13), or the length of the arc M_0M which we shall designate by q, with the retardation or path increment, δ, which is incurred in passing from O by way of M to the field point P, instead of passing from O, by way of M_0, to P. An expression will be derived which holds to a high degree of approximation when the distance of M from M_0 is small compared to a or b. In the figure the distance of M from M_0 is of necessity greatly exaggerated. To the degree of approximation involved, the distances ρ and q may be considered equal. Now:

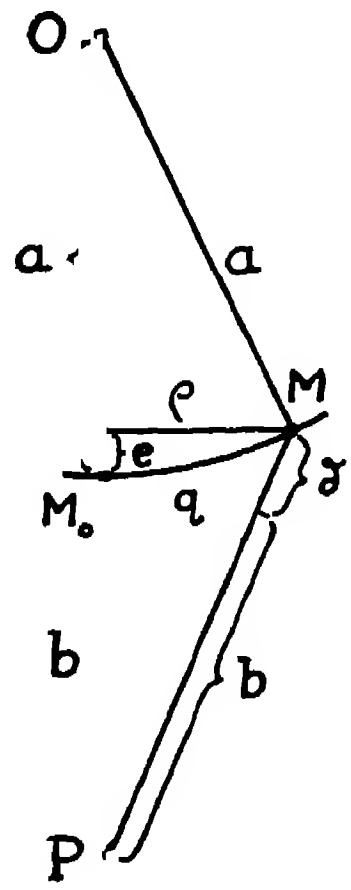


FIG. 13

$$OM^2 = a^2 = (a-e)^2 + \rho^2 \tag{1}$$

where e is the distance indicated. When M lies close to M_0, then e is small compared to ρ or q, hence we may neglect terms containing e^2. Equation (1) thus becomes:

$$e = \frac{\rho^2}{2a}. \tag{2}$$

Denote the distance MP, or $b+\delta$, by r, then:

$$r^2 = (b+e)^2 + \rho^2, \tag{3}$$

whence, again neglecting terms in e^2 and substituting the value of e from equation (2):

$$r^2 = b^2 + \frac{b\rho^2}{a} + \rho^2,$$

whence

$$\rho^2 \frac{(a+b)}{a} = (r+b)(r-b). \tag{4}$$

The difference term on the right $(r-b)$ is the path increment previously designated by the symbol δ, which we shall now substitute for $(r-b)$. The term $(r+b)$ equals $2b+\delta$. Now δ is very small, being of the same order of magnitude as e, hence where δ occurs as an addition to a large term $2b$, we

* *Œuvres complètes*, **1**, 365.

may neglect it and substitute for $(r+b)$ the term $2b$. Confining ourselves to small values of ρ and q, we thus have approximately:

$$\rho = q = \sqrt{\frac{2ab\delta}{(a+b)}}\,. \tag{5}$$

3. Dimensions of Zones of Low Number.—If the point M lies at the outer boundary of the nth zone, then, according to the principle upon which the Fresnel zones are constructed, $\delta = n\lambda/2$. Let us call ρ_n the radius of the outer boundary of the nth zone and q_n the distance measured along the arc to this boundary. When M lies near M_0, δ is small and n will not be very large. Under this condition we may apply equation (5), obtaining:

$$\rho_n = q_n = \sqrt{\frac{nab\lambda}{a+b}}\,. \tag{6}$$

By giving n successively the values 1, 2, 3, etc., we learn that the zones obey the same law which holds for Newton's rings, namely, the radii are proportional to the square roots of the natural numbers: Figure 14 shows the first twenty zones in relative proportion. The zones of odd number, the first, third, fifth, etc., are shown in white and those of even number in black. The figure shows the zones greatly magnified. For example, when the distances a and b are each $1m$ and $\lambda = 5\times10^{-5}$ cm $= 5000$ Å, we deduce from equation (6) that the radius of the first zone is .5 mm.

FIG. 14.—Zone plate

Regarding the areas of the zones: The area of the nth zone, which we shall denote by σ_n, equals the total area comprised within the outer boundary of this zone less the area comprised within the outer boundary of the zone of number $n-1$. Hence, approximately:

$$\sigma_n = \pi\rho_n^2 - \pi\rho_{n-1}^2 = \frac{\pi ab\lambda}{a+b}\,. \tag{7}$$

The expression on the right does not contain n, therefore the areas of the zones of low number are all equal, or rather they may be so thought of for the present. Equation (7) is only approximate. An exact expression, which is however less convenient in form, will be derived in section 13. When the

values of a and b are those given above, the areas of the zones are each less than 1 mm^2.

Equations giving the dimensions of the zones for the special case of plane waves are easily derived. In this case a is equal to infinity and e is equal to zero. Introducing the value $e=0$ into equation (3) and carrying through the same steps as those above given for the general case, we obtain $\rho_n = q_n = \sqrt{bn\lambda}$ and $\sigma_n = \pi b\lambda$, where b is the distance from the plane wave front to the field point under consideration.

4. Circular Aperture; Small Number of Zones.—If the resultant disturbance reaching the given field point P from the first zone is considered positive, then the disturbance reaching this point from the second zone must be considered negative, because the second zone lies on the whole farther from P by the distance $\lambda/2$, and consequently, according to the principle of interference, the disturbance due to the second zone will tend to annul that due to the first zone. The third zone again lies farther from P by a distance $\lambda/2$ and therefore the effect of the third zone tends to annul that due to the second zone and to reinforce that due to the first zone. Therefore the effect of the third zone is to be considered positive. In the same way the effect of the fourth zone will be negative, that of the fifth zone positive, etc. We may thus write for the total effect of the first n zones:

$$S = S_1 - S_2 + S_3 - S_4 \ldots . (-1)^{n-1} S_n , \tag{8}$$

where S_1, S_2, etc., are each positive quantities. The magnitude of the effect of the first zone nearly equals that of the second zone, which in turn nearly equals that of the third zone, etc., because the zones are nearly equal in area and differ only minutely in their distances from P. Suppose now that an opaque screen with a circular aperture in it is erected in the position of the wave front. If the aperture allows the first two zones and only these to pass, the effects of these two should annul each other by interference and there should be darkness at P, whereas if the aperture were *smaller*, allowing only the first zone to pass, then there should be illumination at P. This surprising conclusion is well borne out by experiment. (It should be noted, however, that the conclusion applies only to a point P lying on the axis of the aperture.) When the aperture allows the first three zones to pass, then the effects of two of the zones annul each other and the effect of the third is left outstanding. The illumination is therefore the same as when one zone alone is allowed to pass. When the aperture allows the first four zones to pass there is darkness at P. From the first five zones there is illumination equal to that from one zone alone, from the first six zones, darkness, etc. But, as so far proved, this holds only so long as we confine ourselves to a relatively small number of zones. The general case will be treated later.

5. Zone Plates.—By photographing Figure 14, a so-called "zone plate" may be made. (The zone plate was originated by J. L. Soret.*) The first, third, fifth, etc., zones will be black on the negative and no illumination from them will pass. The effects arising from the second, fourth, sixth, etc., zones will all reinforce each other, consequently, referring to Figure 15, if O is a source point on the axis of the zone plate Z, represented in cross-section, there will be greatly intensified illumination at a point P which is located on the other side of the zone plate, on the axis and at the proper distance from the plate. At no other point in the field plane PP' will there be corresponding reinforcement when O is the only source point, therefore P is an "image" of O formed by the zone plate in much the same manner as a convex lens might form an image. If O' is another source point, not far removed from the axis, then it is obvious that the zone plate will still serve to a high degree of approximation to pass the zones of even number with reference to a field point P', and therefore P' will be an image of O'. Therefore if an object such as the luminous filament of a lamp is placed at OO', then each point of the object will act as a source, and an image of the object, represented by the arrow PP', will be formed in the field plane. This image may be received, for example, upon a sheet of white paper.

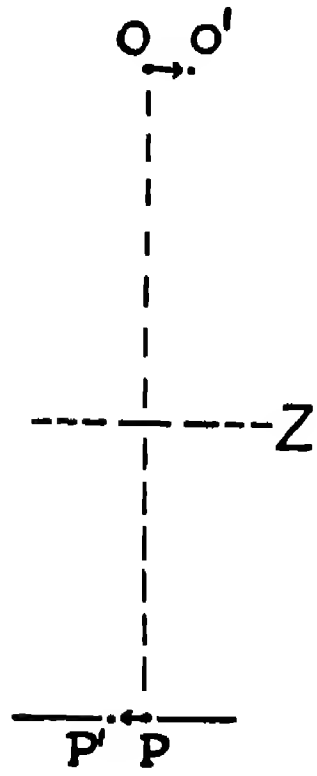


FIG 15

By moving the sheet of paper considerably closer to the zone plate, a second image may be found. The Fresnel zones are now much smaller. This second image occurs at the position in which the Fresnel zones are alternately blocked off and passed in sets of three. The central opaque region of the plate blocks off the first three zones, the first transparent region passes the next three zones, etc. In the sets of three which pass, the effects of two of the zones annul each other and the effect of one is left outstanding. The second image is thus much less intense than the first one. A still fainter third image may sometimes be found at the position in which the zones are blocked off or passed in sets of five. In the sets which pass, the effects of four zones annul each other and the effect of one is left outstanding. Zone plates exhibit the phenomenon of multiple foci.

By using two zone plates of different sizes, having consequently different "focal lengths," one as objective and the other as eyepiece, a *zone-plate telescope* may be made.

Zone plates form not only real images, but also form virtual images.†

* *Compt. rend.*, **80**, 483, 1875, and *Pogg. Ann.*, **156**, 99, 1875.

† See, e.g., Wood, *Physical Optics* (2d ed., Macmillan, 1914), p. 217. The general question of virtual diffraction patterns is briefly discussed in our chap. 5, sec. 8.

In the foregoing discussion it is assumed that the central region of the zone plate is black, the next region transparent, etc. However, we may equally well have the central region transparent, the second black, etc.

The performance of zone plates presents to us forcefully the elements of reality in our theory, but such plates have no practical usefulness as substitutes for lenses. There are several drawbacks to their employment as optical instruments. Among these attention is directed to the fact that the distance at which the image is formed is a function of the wave-length, λ, and therefore a good image is obtained only with monochromatic light. When white light is used the image is bordered by colored fringes.

In making zone plates by photographing a diagram, for example Figure 14, the photographs must be properly reduced in size and must be in good focus. Even so they will not all give good images because of imperfections in the glass of the photographic plate. Examination of the finished zone plates by the method of striae will usually reveal the faults which are present in those which fail to work properly.

Lord Rayleigh* suggested that if instead of blocking off, let us say, the odd zones, the illumination from these were also allowed to pass but with an additional optical path of a half wave-length somehow introduced, then the disturbances from all of the zones would reinforce each other and a more intense image than that given by the ordinary zone plate would result. Several methods of achieving this result have been devised by R. W. Wood.† One of these consists in etching alternate zones to the proper depth on a plate of glass by means of hydrofluoric acid, thus causing the phases of the disturbances from these zones to be reversed. The plates are termed "phase-reversal zone plates."

6. Equations of Wave Motion.—When the Fresnel zones are considered each as a unit, no account is taken of the fact that different points within one zone lie at different distances from the field point. The distance of various points within one zone from the field point varies over a range of $\lambda/2$, hence disturbances from the extreme points arrive at the field point in opposite phases. A thorough treatment should obviously take account of this fact. By way of preparing for such a treatment some simple equations of wave motion will now be discussed.

Denote by ξ the disturbance at time t at a point in space having co-ordinates x, y, and z. This *instantaneous disturbance* may be the displacement of an air particle when we are dealing with sound, or the displacement of an ether particle when we are dealing with light, provided we believe in the

* *Scientific Papers*, **3**, 78.

† *Phil. Mag.*, **45**, 511, 1898, and *Physical Optics*, p. 40.

ether, or the instantaneous electric or magnetic intensity. We do not need to specify the nature of ξ. Whatever it may be, let us write:

$$\xi = B \cos 2\pi \left(\frac{t}{\tau} - \frac{x}{\lambda}\right) \tag{9}$$

and regard B, τ, and λ as constants. It will be shown that this expression represents plane waves traveling along the X-axis of co-ordinates in the positive direction. It is at once evident that if the equation represents wave motion at all the waves must be plane. This is true because y and z do not appear in the expression, and therefore the value of ξ is independent of y and z—or, in other words, has a constant value over any plane perpendicular to the X-axis. Continuing, at time $t_0 = 0$ we have $\xi = B \cos [-2\pi x/\lambda]$, and giving x successively the values 0, $\lambda/8$, $\lambda/4$, $3\lambda/8$, etc., the circles on the solid curve of Figure 16 are obtained. At time $t = \tau/8$,

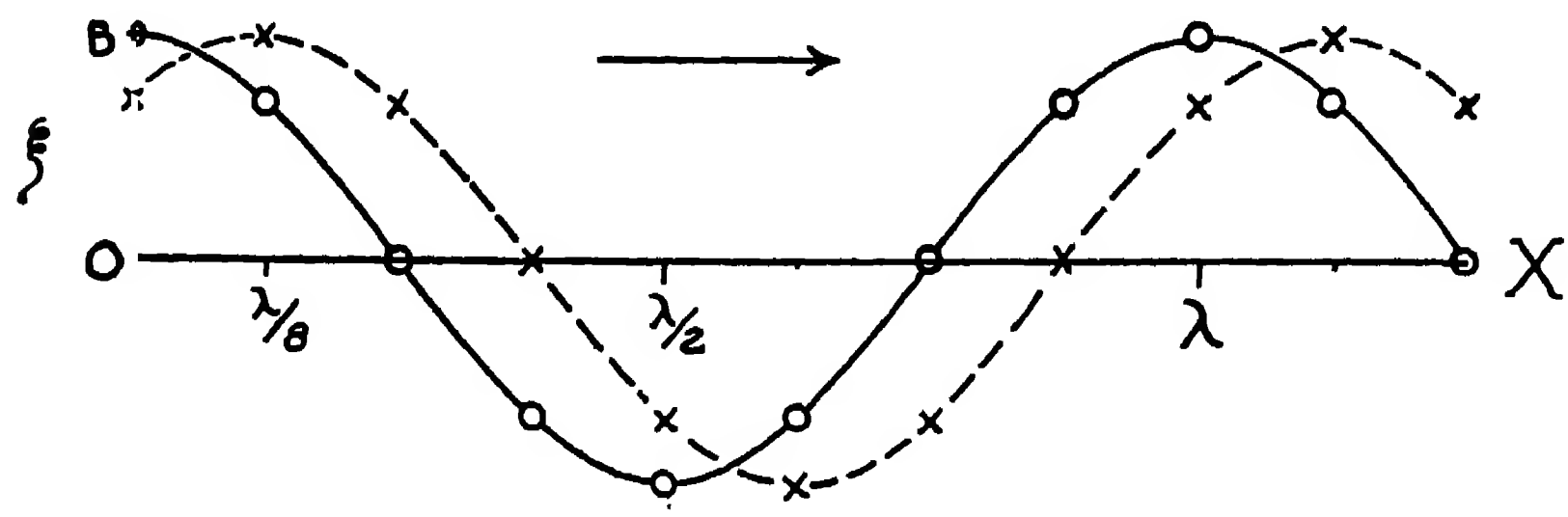


FIG. 16

we have $\xi = B \cos 2\pi[1/8 - x/\lambda]$, and giving x the same succession of values as before the crosses on the broken curve are obtained. The broken curve is the same as the solid curve, only displaced in the positive X-direction by a distance $x = \lambda/8$. At time $t = \tau/4$ the curve would be displaced as far again. Equation (9) thus represents a train of plane waves traveling to the right. The amplitude of the waves is B. This is constant throughout all space when the waves are not absorbed in their progress through the medium. When t changes by an amount equal to τ, the value of ξ repeats itself and all of the derivatives of ξ repeat themselves, hence τ is the period. When x changes by an amount λ, the value of ξ likewise repeats itself and so do the values of the derivatives, hence λ is the wave-length.

Spherical waves may be represented by replacing x, the distance measured along an axis, by the radial distance of the wave front M_0M (Fig. 17) from the source O. This radial distance will be denoted by the symbol a which is to be considered as a variable. Account must also be taken of the

fact that spherical waves decrease in amplitude with increasing distance from the source. According to a familiar relation:

$$\frac{I_1}{I_2}=\frac{a_2^2}{a_1^2}, \tag{10}$$

where I_1 is the intensity at distance a_1 and I_2 is the intensity at distance a_2. The intensity of a wave is proportional to the square of its amplitude, hence the amplitude of spherical waves varies as the simple inverse of the distance from the source. If we call A the amplitude at unit distance from O, then the amplitude at distance a is A/a. Hence a simple harmonic train of spherical waves may be represented by the expression:

$$\xi=\frac{A}{a}\cos 2\pi\left(\frac{t}{\tau}-\frac{a}{\lambda}\right). \tag{11}$$

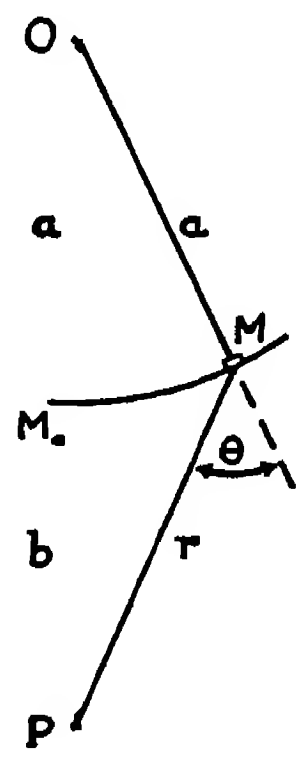


FIG. 17

7. The Contribution from an Elementary Area.—Let $d\sigma$ represent the area of a small element taken at M (Fig. 17) on the wave front M_0M. This element is to be regarded as infinitesimal—that is, its length or breadth is to be regarded as vanishingly small compared even to the small width of a Fresnel zone. The element is, according to Huygens' principle, the source of a spherical wavelet, and an expression is to be found for the effect of this wavelet at the field point P. The amplitude of the issuing wavelet is proportional to A/a, which is the amplitude of the original wave which obtains on the wave front M_0M. The amplitude of the wavelet must also be taken as proportional to the area of the element under consideration, hence the expression for the amplitude of the wavelets must contain the factor $d\sigma$, as well as the aforementioned factor A/a. Moreover, the amplitude of the wavelet at a given point P varies inversely as the distance of P from the element of area. This distance must therefore occur in the denominator; it will be denoted by r. The amplitude due to an infinitesimal element of area is itself infinitesimal. Hence the amplitude at P due to the element $d\sigma$ may be written:

$$dQ=\frac{k\,A\,d\sigma}{a\,r}, \tag{12}$$

where k is an undetermined factor which is not a function of the values of A, a, $d\sigma$, or r, but may depend upon the angle θ between the normal to the element of area and the direction in which the point P lies.

The phase of the disturbance at P is determined by the phase of the dis-

turbance on the wave front M_0M, and also by the distance r which the wavelets must traverse to reach P. The phase which obtains on the wave front M_0M is $2\pi(t/\tau-a/\lambda)$. The phase at P is retarded relative to this by a phase angle $2\pi r/\lambda$. Hence the phase at P is $2\pi[(t/\tau-a/\lambda)-r/\lambda]$. The instantaneous disturbance at P arising from the element $d\sigma$ is thus:

$$ds=\frac{k\,A\,d\sigma}{a\,r}\cos 2\pi\left[\left(\frac{t}{\tau}-\frac{a}{\lambda}\right)-\frac{r}{\lambda}\right]. \tag{13}$$

The disturbance ds varies with the time t and constitutes an infinitesimal simple harmonic disturbance. It is the contribution from a single element of area of the wave front. To it must be added the contributions from other elements, taking account of the phases in which these contributions arrive.

8. The Summation of Simple Harmonic Disturbances.—Referring to Figure 18, let p and q represent two points which are each executing simple harmonic motion along the horizontal line indicated. Let the periods of the two motions be the same, but let the amplitudes and phases be different. Let OP and OQ be radii of the reference circles drawn for the motions p and q, respectively. The lengths OP and OQ thus represent the amplitudes of the two motions. The points p and q may respectively be regarded as the projections of the two points P and Q moving uniformly, with a common angular velocity, in the circles of radii OP and OQ. Complete the parallelogram formed by the radii OP and OQ and project the point R upon ON, obtaining r. The instantaneous displacement of r measured from O is $Or=Op+pr=Op+Oq$. The point r thus executes a motion which is the sum of the motions of p and q. Since the motions of p and q have the same period, the angle between the vectors OP and OQ remains constant as time progresses and hence the parallelogram rotates with unaltered form and the point R moves uniformly in a circle of radius OR. Thus the point r also executes simple harmonic motion of the same period, and we arrive at the conclusion that the sum of any two simple harmonic motions along the same line and of the same period is itself a simple harmonic motion of this period. The simple harmonic motions of the points p and q are represented in amplitude and phase by the radii OP and OQ. The length of each radius measures the amplitude of the corresponding motion and the position of each radius determines the phase of each motion. Let us measure all phase angles clockwise from ON as zero line. The summational motion, r, is now represented by the radius OR, whose length represents the amplitude of the motion of r, and the angle from ON, clockwise, to OR, represents the phase of the motion of r. Thus each radius in the diagram is equivalent to a vector. The two original simple harmonic motions may be represented by vectors, just as forces are represented by vectors, and the resultant simple har-

monic motion is then represented by a vector which is the sum of the two original vectors. This sum may be found by graphical addition or the addition may be carried out analytically. In the latter case the resultant amplitude is represented by one expression, and the phase by another, these expressions corresponding respectively to expressions representing one the magnitude and the other the direction of a resultant force.

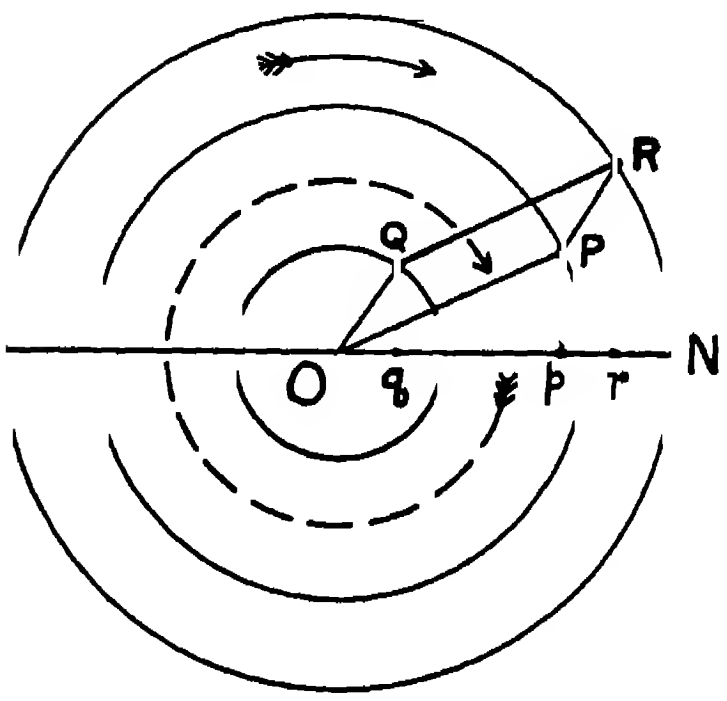


FIG. 18

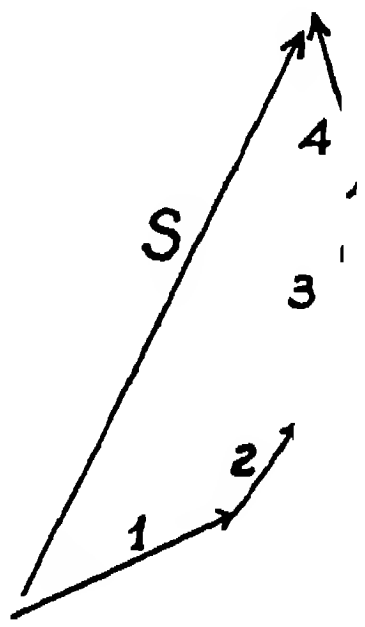


FIG. 19

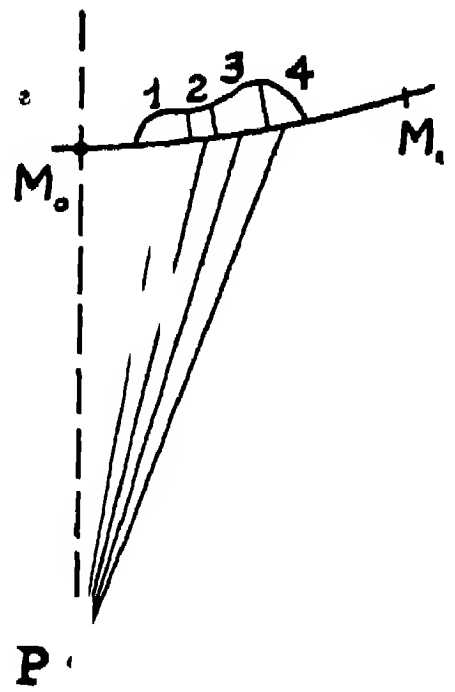


FIG. 20

Since two simple harmonic motions when added yield a simple harmonic motion, this sum may be added to a third simple harmonic motion to yield still another simple harmonic motion, etc. Thus referring to Figure 19, suppose the vectors 1, 2, 3, and 4 represent each a simple harmonic motion, then the sum of these motions is represented in amplitude and phase by the sum of the vectors, namely, by the vector S.

The four vectors may be considered to represent four simple harmonic disturbances reaching the field point P (Fig. 20) from the elements of area numbered 1, 2, 3, and 4, standing upright, on the wave front M_0M_1, each element being supposedly very small. For the sake of definiteness let us think of the point M_1 as marking the outer edge of the first Fresnel zone.*

Denote the areas of the elements by $(\Delta\sigma)_1$, $(\Delta\sigma)_2$, etc., and the distances of their centers from P by r_1, r_2, etc. In a given specific case we may determine the simple harmonic disturbances which reach P by applying equation (13) for each of the elements provided the elements are sufficiently small. The lengths of the vectors 1, 2, etc., in Figure 19 represent the am-

* The above-suggested representation involves the supposition that the simple harmonic disturbances reaching the field point P from the elements of area 1, 2, 3, etc., all lie in the same straight line. One might raise the objection that the disturbances will not all lie along the same line because the elements of area, as regarded from the field point P, lie in different directions. But when the elements of area are minute and contiguous, the differences in direction will be minute. If the reader will but proceed with his study of the theory at hand, he will realize more and more as he proceeds that the objection in question is only an apparent one and does not really compromise the validity of the theory.

plitudes of the disturbances reaching P and these amplitudes are proportional to $(\Delta\sigma)_1/r_1$, $(\Delta\sigma)_2/r_2$, etc. The directions of the vectors are determined by the phases in which the disturbances arrive at P, and these phases are in turn determined by the distances r_1, r_2, etc.

The vectors 1, 2, 3, etc., form an open polygon and the closing vector S of this polygon represents in amplitude and phase the resultant disturbance at P due to the entire area under consideration—or, rather, it represents this disturbance approximately. By dividing the elements of area, let us say each into two parts, we would obtain a closer approximation. There would then be twice as many vectors each half as long as before and making half as great angles with each other on the average. By subdividing the elements indefinitely—that is, by passing to the limit—we obtain a smooth curve (Fig. 21), and the closing vector S of this curve represents accurately the simple harmonic disturbance at P due to the entire area.

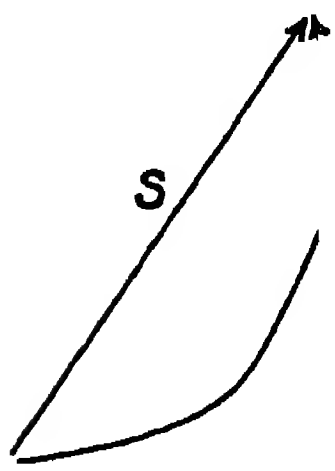


FIG. 21

The process above outlined is a graphical method of performing an integration. We shall learn in following sections how the method is to be applied.

Diagrams such as Figures 19 and 21 will be referred to as *vibration diagrams*, those of the type of Figure 19 as vibration *polygons*, and those of the type of Figure 21 as vibration *curves*.

9. Elementary Zones.—Referring to Figure 22, suppose that M is a point on a wave front which lies within the first Fresnel zone, the outer boundary of this zone being indicated by M_1. Denote the distance MP by r and the angle MOP by ϕ. Suppose M' is a point at an infinitesimal distance $dq = MM'$, measured along the wave front. Denote the distance $M'P$ by $r+dr$ and the angle $M'OP$ by $\phi+d\phi$. If the points M and M' are rotated about OP as axis, they will outline an elementary zone of infinitesimal width upon the wave front. All portions of this infinitesimal zone bear the same relation to the field point P, hence we may treat this zone as a single element of area in applying the theory previously given. The width of the zone is $MM' = dq = a\, d\phi$. Denoting the radius of the elementary zone by ρ, the circumference is $2\pi\rho = 2\pi a \sin\phi$. The area of the zone is the product of its width and its circumference, hence the area is:

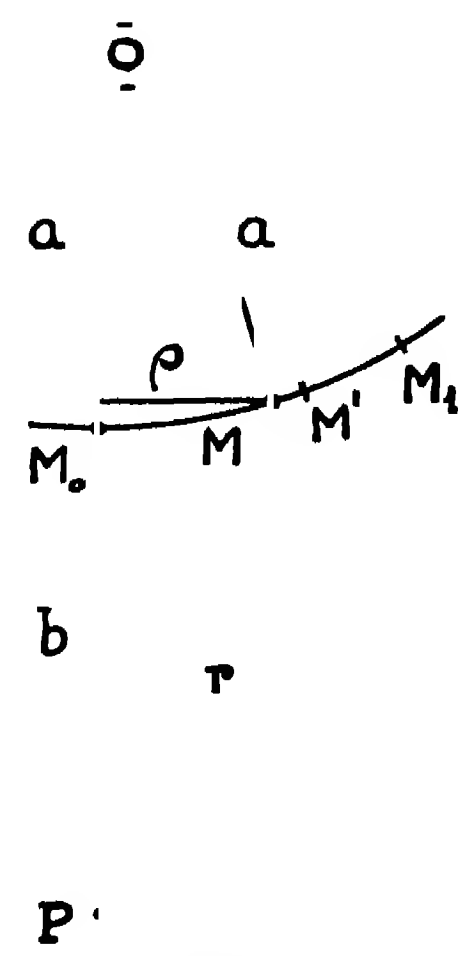


FIG. 22

(14) $$d\sigma = 2\pi\rho dq = 2\pi\rho a\, d\phi = 2\pi a^2 \sin\phi\, d\phi\,.$$

By a familiar trigonometric relation:

(15) $$r^2 = a^2 + (a+b)^2 - 2a(a+b)\cos\phi\,.$$

Differentiating:

(16) $$r\,dr = a(a+b)\sin\phi\,d\phi.$$

Eliminating $\sin\phi\,d\phi$ from equations (14) and (16):

(17) $$d\sigma = \frac{2\pi a\,r\,dr}{(a+b)}\,.$$

This value of $d\sigma$ may be substituted in equation (13), yielding:

(18) $$ds = \frac{2\pi kA\,dr}{(a+b)}\cos 2\pi\left(\frac{t}{\tau} - \frac{a}{\lambda} - \frac{r}{\lambda}\right).$$

Here ds is the instantaneous value of an infinitesimal simple harmonic motion; ds represents the instantaneous contribution toward the disturbance at the field point P which arises from a zone of infinitesimal area $d\sigma$ lying at a distance r from the field point. Now in deriving equation (13), the amplitude of the contribution from an area $d\sigma$ was taken as proportional to $d\sigma$ and inversely proportional to the distance r of the area from the field point. However, we learn from equation (17) that $d\sigma$ for an elementary zone is itself directly proportional to r. Hence in eliminating $d\sigma$ as a term appearing explicitly, as was done in obtaining equation (18), r drops out of the amplitude term. But dr is still present in this term—the amplitude is directly proportional to dr.

Let us now construct a large number, m, of elementary zones of minute but finite width. These m minute zones, taken collectively, shall comprise the first Fresnel zone and shall be so constructed that the increment in the variable r in passing from one boundary to the next shall be always the same, namely, $\Delta r = (\lambda/2)/m$. Thus the boundary of the central minute zone shall lie at a distance $r_1' = b + (\lambda/2)/m$ from P, the outer boundary of the second one shall lie at a distance $r_2' = b + \lambda/m$ from P, etc., until we reach the outer boundary of the mth minute zone which coincides with the boundary of the first Fresnel zone and lies at the distance $r_m' = b + m\Delta r = b + \lambda/2$ from P. By inserting Δr in equation (18) in place of dr, we may apply this equation approximately to the minute zones. These will send contributions of equal amplitude to P because all terms in the expression for the amplitude are constant including dr which is replaced by the constant Δr. (We must make a reservation in regard to the constancy of k. But any variation which

k may have will be overlooked for the moment.) The vibration polygon for the minute zones is thus, aside from any effect introduced by the variation of k, a polygon of m vectors of equal length, where m is to be thought of as a large number. In making drawings m cannot conveniently be taken very large, and it will be chosen as equal to 12. In accordance with this choice, Figure 23a represents the polygon. The vectors are of equal length. Moreover, they make equal angles with each other. This is true because r appears linearly in the phase term of equation (18), and since r increases by equal amounts, Δr, in passing from each minute zone to the next, the phase is retarded by equal amounts in passing from one such zone to the next. Furthermore, the first vector and the mth one point in approximately opposite directions since values of r corresponding to the first and the mth minute zone differ by $\lambda/2$ approximately, and hence the phases differ approximately by π. The polygon is both *equilateral* and *equiangular* and may consequently be inscribed in a semicircle. The closing vector of the polygon, S_1, represents in amplitude and phase the illumination due to the first Fresnel zone. By considering m to increase indefinitely and Δr to decrease correspondingly in such a manner that the relation $m\Delta r = \lambda/2$ is always satisfied, we may pass from the vibration polygon of Figure 23a to a corresponding vibration curve which would be accurately a semicircle, provided k were absolutely constant.

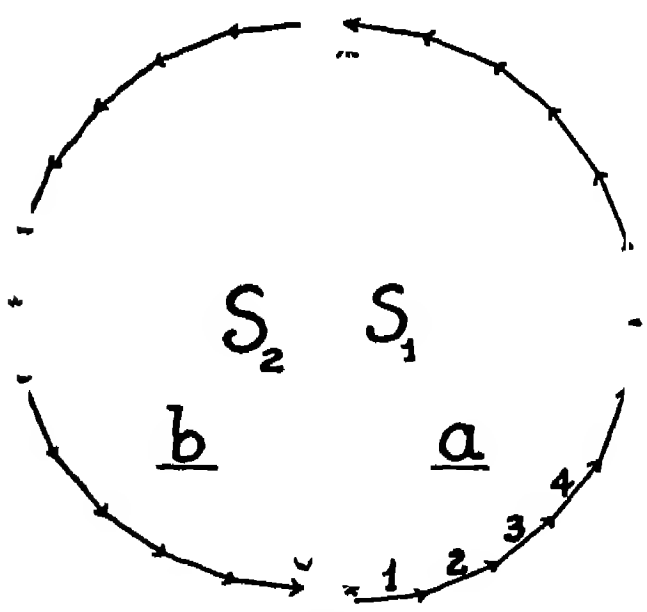


FIG. 23

10. The Inclination Factor.—Let M (Fig. 24) represent an element of the original wave front and let ww represent the Huygens wavelet to which this element gives rise. Fresnel made the reasonable assumption that the wavelet has a maximum amplitude in the direction MN normal to the original wave front and that the amplitude of the wavelet decreases as the angle of inclination θ increases. This decrease of amplitude upon receding from the normal is indicated in the figure by the decrease in the width of the heavy solid circle which represents the wavelet. The aforementioned factor k varies in proportion to the variation of amplitude with inclination and is hence known as the *inclination* factor, or *obliquity* factor. It is not necessary to know the law according to which the variation of amplitude with inclination takes place. In fact, in so far as the requirements of diffraction theory are concerned, the amplitude does not even need to decrease continually with increasing inclination; there might be intermediate stages in which the sense of the variation is reversed. However, leaving

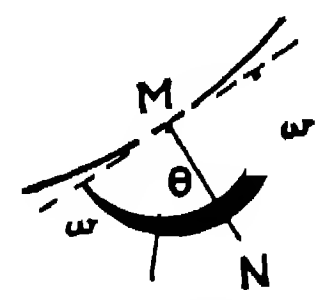


FIG. 24

this latter question aside, we shall, following Fresnel, suppose that k has a maximum value at zero inclination and decreases continuously as θ increases, without assuming any specific law for the variation of k. Fresnel did, however, assume that the amplitude of the wavelet decreases to the value zero in the direction tangent to the original wave front and is zero for all backward directions. We shall for the time being follow Fresnel in this assumption also. This assumption is closely associated with the question of why the original wave is propagated only forward, and not backward—that is, why there is no so-called "back wave." In fact, various questions arise in this connection, but it would be premature to consider these at present. These questions are considered in chapter 7, sections 25–27.

Returning now to pick up the thread of previous discussion: Owing to the decrease in the inclination factor k in passing from the normal outward, the vectors 2, 3, 4, etc., of Figure 23a are each one ever so little shorter than the preceding one. The angles which the vectors make with each other are however accurately equal—for reasons which have been pointed out. As a result of these two circumstances considered in conjunction, when we go over to the corresponding vibration *curve*, and proceed along this curve, the curvature of the curve becomes greater as we proceed. But the curvature becomes greater by only an extremely slight amount.

Let us now consider the construction of the minute zones to be carried on through the second Fresnel zone. We again obtain a vibration polygon (Fig. 23b). The closing vector S_2 of this polygon is directed oppositely to S_1 and is very slightly shorter than S_1. By passing to the vibration curve for the second Fresnel zone and adding this to the vibration curve for the first one, we obtain Figure 25. The closing vector of this curve is not shown, because its length is nearly zero. The almost zero length of this vector repre-

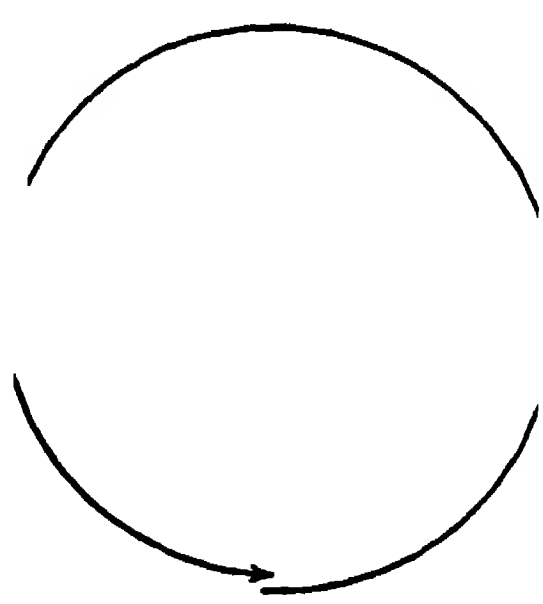


FIG. 25

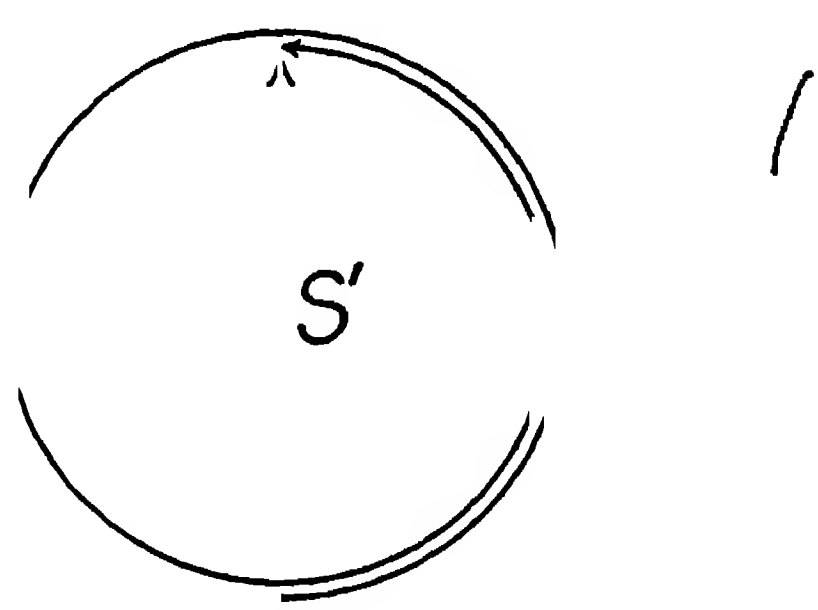


FIG. 26

sents the total effect of the first and second Fresnel zones taken together—that is, the first two zones, practically speaking, yield darkness, as we have already learned. The vibration curve for the third Fresnel zone is an almost

exact semicircle lying just inside of the vibration curve for the first zone (see Fig. 26). The closing vector S' of the curve for the first three zones taken together represents the resultant illumination from these zones. The length of this closing vector is practically equal to the length of the closing vector for one zone alone.

11. Circular Aperture; Any Number of Zones.—As we continue the construction of the vibration curve for numerous additional zones, the curve winds up appreciably. The resulting effect is represented by the spirals of Figure 27, only the rate of diminution in radius which is shown in the figure is far too rapid. The spiral on the left is for a circular aperture which allows

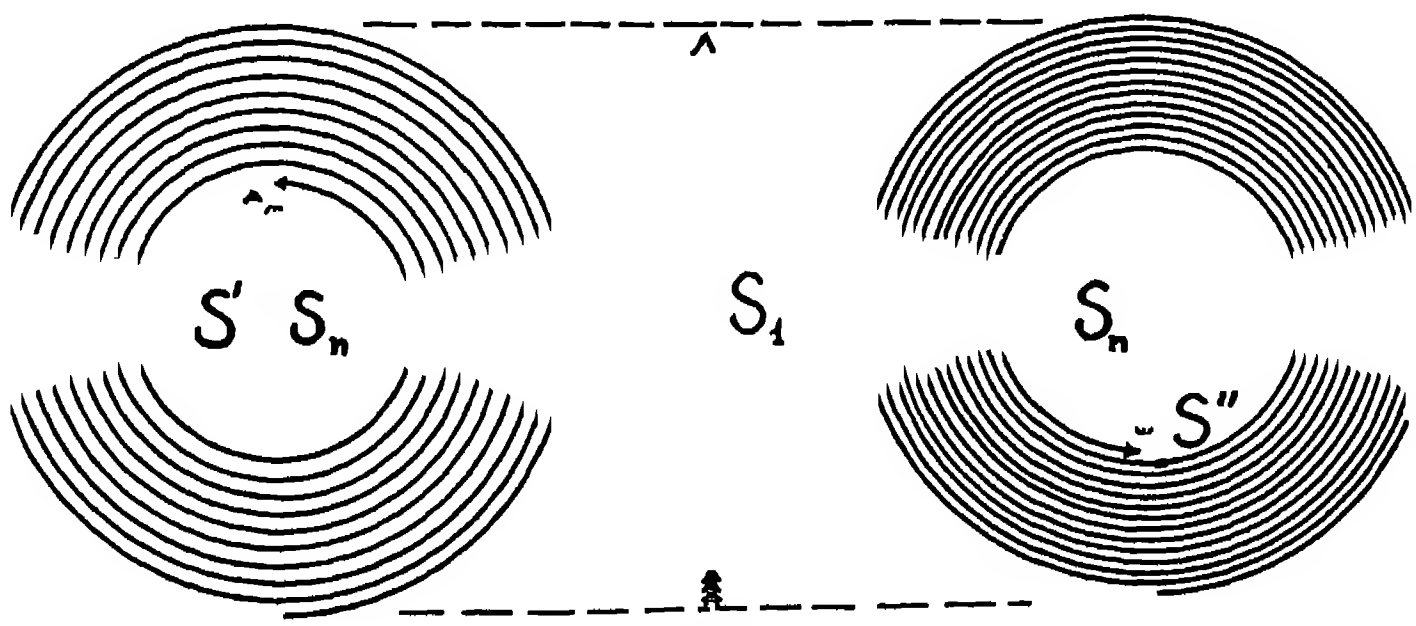


FIG. 27

an *odd* number of zones to pass. The closing vector, S', of the spiral is a measure of the amplitude of the resultant illumination. Denote the given odd number of zones which pass by n, then:

$$S' = \tfrac{1}{2}S_1 + \tfrac{1}{2}S_n\,, \tag{19}$$

where S_1 and S_n are the closing vectors of the first and nth zones, respectively.

When an even number of zones passes the aperture, the resulting vibration curve may be represented by the spiral on the right of the figure. Again denoting the number of zones which pass by n, the length of the closing vector, measuring the amplitude of the resulting illumination, is:

$$S'' = \tfrac{1}{2}S_1 - \tfrac{1}{2}S_n\,. \tag{20}$$

Equations (19) and (20) apply for either a large or a small number of zones. They include as a special case the results previously arrived at when the number of zones is small. In this case S_n nearly equals S_1, and equation (19) tells us that when n is a small odd number the total effect is practically equal to that of a single zone, let us say of the first zone. Equation (20) tells us that when n is a small even number the total effect is practically zero.

The proof of equations (19) and (20) was first given by A. Schuster by an analytical method.* The truth of these equations is self-evident from consideration of the vibration diagrams. They give the illumination only at a field point P lying on the axis of the aperture. The illumination which obtains in the remainder of the field will be discussed in chapter 4.

12. The Complete Wave Front.—The vibration diagram for the complete wave front is now easily arrived at. As the effect of zone upon zone is added, the radius vector, r, from the field point to the wave front, eventually approaches tangency to the wave front. According to Fresnel's postulates, the value of the inclination factor approaches zero, and hence the magnitude of the effect of the last, the nth zone, approaches zero. As shown in Figure 28, the vibration curve winds up about the center of the nearly circular arcs formed by its constituent convolutions. The length of the closing vector S drawn from the beginning of the vibration curve to this center is a measure of the amplitude of the illumination due to the complete wave front. The vector S equals $\frac{1}{2}S_1$. We thus arrive at the important conclusion that the amplitude of the illumination due to the whole wave front is equal to *one-half* of the *amplitude due to the first zone.* Several remarks are in place regarding this conclusion: One-half of the amplitude due to the first zone is here in question, and *not* the amplitude due to half of the first zone, which is different. The latter is represented by the length of the vector S''' (Fig. 29). Second, instead of specifying the first zone, we might with almost equal justification say any zone of low number, because the effects of the zones of low number are all nearly equal in magnitude. Further, the complete wave front yields an amount of illumination which will from now on serve us as a standard of comparison. When a diffracting object is used—for example, an opaque screen with a circular aperture which allows only the central zone to pass—we shall compare the illumination then obtained with the illumination obtained from the complete wave front rather than vice versa. We shall henceforth speak of the illumination due to the complete wave front as "normal illumination." This term will be used to denote the illumination obtained when the wave front is entirely unobstructed, or when such obstacles as may be present do not modify the illumination at the field point

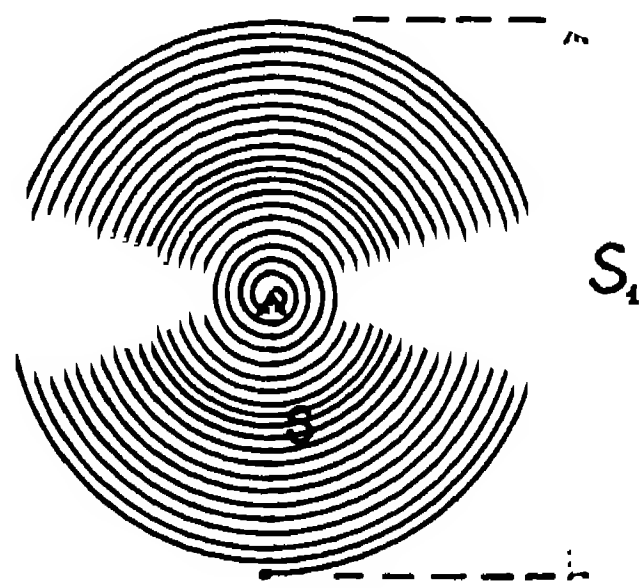


FIG. 28

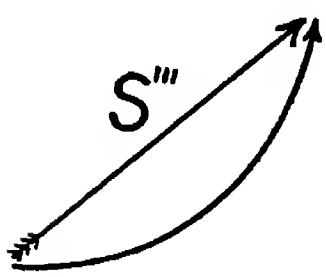


FIG. 29

* For the analytical treatment see, e.g., Schuster and Nicholson, *The Theory of Optics* (London: Edward Arnold), p. 89; or Wood, *Physical Optics* (2d ed.), p. 34.

under consideration. Finally, the closing vector of a vibration diagram is a measure of a resultant *amplitude*, not of an *intensity* of illumination. The intensity varies as the square of the amplitude. Therefore, speaking in terms of intensity, and in terms of normal illumination as a standard, the intensity due to the first zone is *four* times that of normal illumination.

Be it pointed out also in the present connection that Fresnel's theory leads to an inconsistency in regard to the phase with which the resultant disturbance is considered to arrive. Referring again to Figure 28, the phase of the resultant disturbance from the complete wave front is represented by the vector S—that is, by a vector pointing vertically upward. This means that the disturbance from the complete wave front is represented as being retarded by one-quarter of a period, by $\pi/2$ or by $\lambda/4$, with reference to the disturbance coming from the pole of the wave, which is represented by an element taken at the beginning of the vibration curve—or, in other words, by an elementary vector pointing horizontally to the right. Actually, however, the disturbance from the complete wave front arrives in the same phase as the disturbance from the pole. Hence the theory leads to an inconsistency in regard to the resultant phase. The matter here involved will be further discussed in chapter 7, sections 25–27.

13. The Areas of the Zones.—An exact expression for the area, σ_n, of the nth zone may be derived by integrating equation (17) between the limits r_{n-1} and r_n. We have:

$$\sigma_n = \frac{2\pi a}{a+b}\int_{r_{n-1}}^{r_n} r\,dr = \frac{\pi a}{a+b}(r_n^2 - r_{n-1}^2) \tag{21}$$

$$= \frac{\pi a}{a+b}(r_n - r_{n-1})(r_n + r_{n-1}) .$$

Now $(r_n - r_{n-1})$ is, by the principle upon which the zones are constructed, equal to $\lambda/2$. Further, for $(r_n + r_{n-1})$ let us substitute $2\bar{r}_n$ where $\bar{r}_n$ will then denote the mean of the radii vectores from the field point under consideration respectively to the inner and outer edges of the nth zone. We then have for the area of the nth zone:

$$\sigma_n = \frac{\pi a \lambda \bar{r}_n}{a+b} . \tag{21'}$$

The area of a zone thus increases in proportion to $\bar{r}_n$, the distance of the zone from the field point. This increase in area, if it were the only factor involved, would cause the effects of successive zones to increase. But the increase in distance is a factor which in itself causes the effects of successive zones to decrease. Since the area increases proportionately to the distance,

the two tendencies just offset each other and *the magnitudes of the effects of the various zones would all be equal were it not for the decrease in the value of the inclination factor as we proceed to zones of higher number.*

The *absolute* increase in area in passing from one zone to the next, measured let us say in cm^2, is:

$$\sigma_{n+1}-\sigma_n=\frac{\pi a\lambda}{a+b}(r_{n+1}-r_n)=\frac{\pi a\lambda^2}{2(a+b)}\,. \tag{22}$$

This increase is thus a constant independent of n. And it follows further that the *proportionate* increase in area in passing from one zone of *low* number to the next is *greater* than the proportionate increase in area in passing from one zone of *high* number to the next.*

One frequently encounters wrong statements regarding the manner in which the above mentioned quantities vary, and regarding the reasons for their variation. These wrong statements are usually made in the form of conclusions drawn from approximate relationships similar to those represented by our equations (6) and (7). The relationships themselves are correct as approximations, but they are often subsequently used in a manner which fails to take proper account of their approximate nature, and incorrect conclusions are drawn. The fallacies involved in drawing the conclusions are likely to escape the immature reader and the stated conclusions then mislead him.

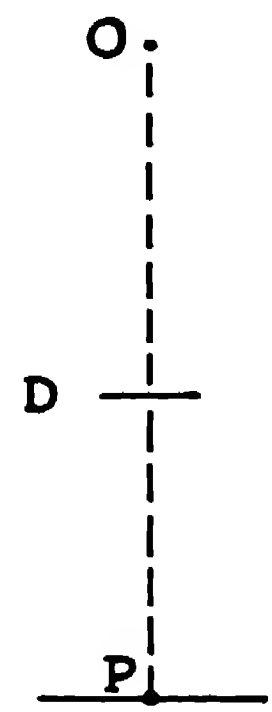


FIG. 30

14. Circular Obstacle.—Referring to Figure 30, suppose that O is a point source of light and D a circular disk mounted with its axis passing through the source O. What will be the illumination at a field point P lying on the opposite side of the disk from O and also on the axis? The disk will obstruct a certain number of zones entirely—let us call this number $n-1$. In general, the disk will also obstruct the inner portion of the next, the nth zone, and will allow the remainder of this zone to pass. All zones beyond the nth will pass completely. The vibration spiral consequently begins at some point f (Fig. 31), and from this point on it winds up exactly as though the disk were absent. The length of the closing vector, S, is a measure of the amplitude of illumination at the field point under consideration. If the num-

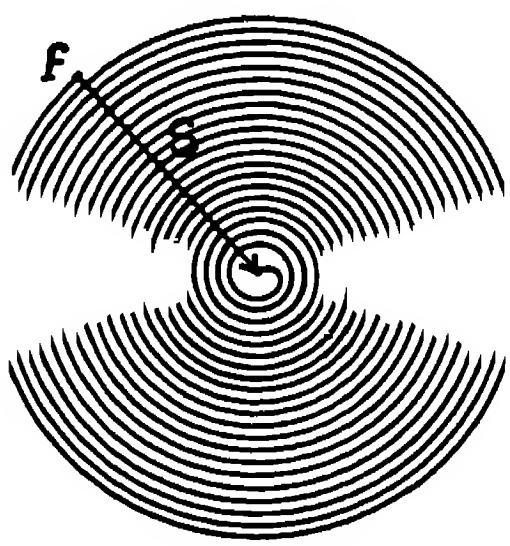


FIG. 31

*It is readily deduced from equations (21′) and (22) that when a and b are each $1m$ and $\lambda=5000$ Å, the proportionate change in area of successive zones of low number is only 2.5×10^{-7}.

ber n is not very large, the vector S will have practically the same length as the vector which would represent normal illumination. That is, the illumination at any point on the axis of the disk should be practically the same as though the disk were absent—a surprising result which is, however, borne out by experiment (see the photograph of Fig. 32, left side).

A different way of stating the explanation is this: We may construct zones, not from the pole of the wave, but from the edge of the disk outward. The sum of the effects of all of the zones outside of the disk will then be equal to one-half of the effect of the first zone which passes the disk.

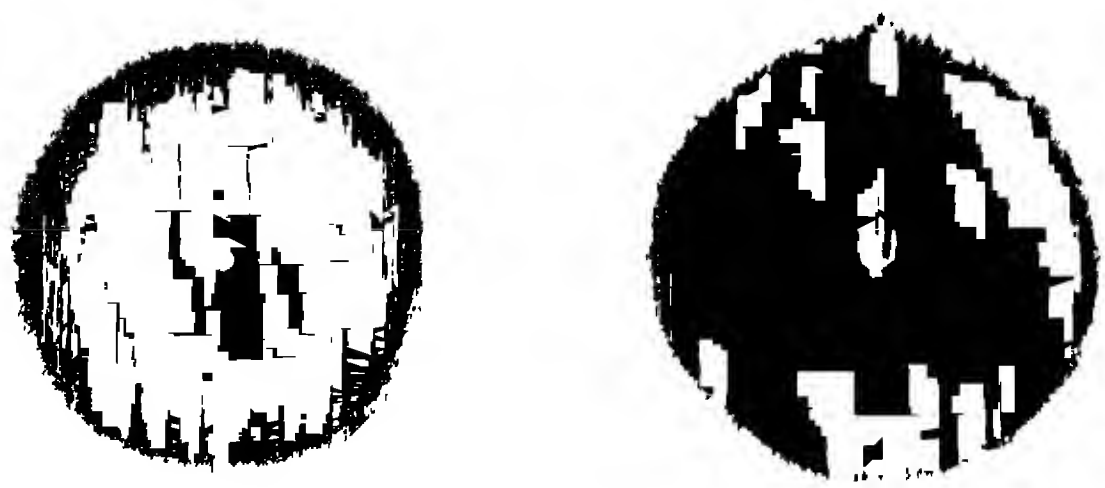


FIG. 32

When the disk is at all large, the experiment becomes difficult to perform. The reasons for this are discussed in chapter 4, section 7. Suffice it for the present to state that when the shadow is to be received upon a sheet of white paper, an intense primary source of light such as an arc should be used. The spot, however, can be readily found even when using a less intense source if the shadow is examined with a lens or eyepiece (without using a diffusing screen, in the manner explained in chap. 1, sec. 4). The spot can also be easily photographed. If the eye is placed on the axis of the disk and the disk is viewed directly, a ring of light appears completely around the edge.*

In place of the disk, a sphere may be used as the diffracting object. There is no distinction to be drawn between the case in which a disk is used and that in which a sphere is used—at least, not in so far as we are at present concerned.

The conclusion that a bright spot should be found at the center of the shadow of a disk was deduced from Fresnel's theory in 1818 by S. D. Poisson who was a member of the committee of the French Academy to which Fresnel's prize memoir on diffraction was referred for review. Poisson was hostile to Fresnel's views and is said to have considered this conclusion as contrary to experiment and hence as refuting the theory. He communicated the deduction first to F. Arago, who was also a member of the committee

* The conditions under which a diffracting edge appears luminous and the reasons for this are discussed in chap. 7, secs. 10 and 11.

and was very friendly toward Fresnel. Thereupon Arago performed the experiment and found the surprising conclusion confirmed.* However, the observation had already been made about a century earlier, and had been forgotten.†

In performing the experiment a small coin with a smooth edge serves well as the disk. It may be supported by two threads. A coin with a milled edge must not be used. A steel or bronze ball through which a hole has been drilled to allow the passage of a thread is easier to mount, although otherwise not superior to a disk. The distances from the source to the diffracting object and thence to the field plane should each be several meters. When it is inconvenient to have these distances so great, a smaller diffracting object —one of three or four millimeters' diameter—may be used. By cementing the object to a plate of glass, the use of supporting threads may be obviated, but the glass must be of good quality.

If O (Fig. 30) is the only source point and the disk or sphere does not obstruct all too small a number of zones, then P is the only point in the shadow which receives appreciable illumination. That is, we may regard P as an "image" of O. Further, if another source point exists, let us say, just to the right of O, an image of this other source point will be formed just to the left of P. Or if the source is a small luminous object, an image of this object will be formed in somewhat the same manner as an image is formed by a lens or a zone plate. Thus it is possible to photograph an object "around" a disk or a ball. In Figure 32, on the right is a reproduction of a photograph so made, the object consisting in this case of a minute photographic transparency.

15. Rectilinear Propagation.—To arrive at the law of rectilinear propagation (which holds only grossly), it will be desirable to begin by considering the vibration curve for a *section* of a zone, the remainder of the zone being

* See Fresnel, *op. cit.*, **1**, xlii, 236, 245, and 365.

† In 1723 Maraldi observed the central bright spot in the shadow of each of several small spheres of different diameters and at various distances (see *Mem. Acad. Roy. des Sci.* [1723], pp. 111 ff.; in particular, p. 139).

E. Verdet (*Leçons d'optique physique*, **1**, 250) stated that the astronomer Jos. Nic. Delisle was the first to make the observation in question, and in support of this statement Verdet cited: *Mem. Acad. Roy. des Sci.* (1715), p. 166. Verdet was a distinguished scholar and teacher and his statement has been widely quoted. However, in the paper referred to, Delisle records observation of the fringes *surrounding* the shadow of a disk but makes no mention of finding a bright spot in the center of the shadow. Nor does he mention finding this spot in any of several other papers. In particular, he fails to mention it in a long paper containing many interesting observations which was presented to the Académie Royale in 1717 but remained unpublished until 1738 when Delisle incorporated it in his St. Petersburg *Memoirs*. Why Verdet should have attributed the first observation to Delisle is somewhat perplexing.

Related observations by De la Hire, though not important, are to be found *Mem. Acad. Roy. des Sci.* (1715), p. 161.

supposedly blocked off. Figure 33 represents sections of two zones, one of low number, *ll*, lying near the pole of the wave, M_0, and the other, one of high number, *hh*, lying far out. In Figure 34, the nearly exact semicircle *c* may be considered to represent the vibration curve which would obtain for one of these zones, either *ll* or *hh*, if the zone in question were effective in its entirety, and the proportionately smaller semicircle *e* may be considered to represent the vibration curve for the effective section. When a section of a zone terminates with square ends as illustrated, the vibration curve for the effective section will be similar to that for the whole zone, only it will be smaller.

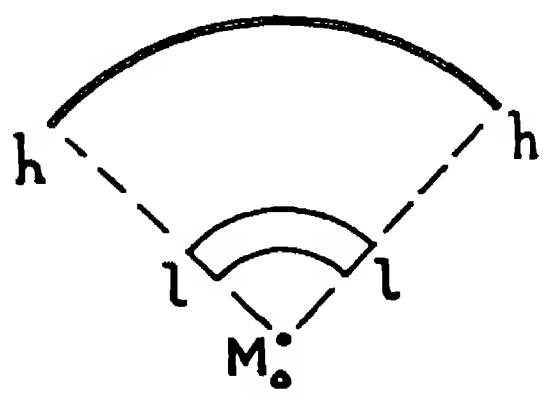


FIG. 33

FIG. 34

Further, a section of a zone of *high* number will yield a vibration curve which is practically a semicircle even when the section does not terminate with square ends. This follows from the fact that zones of high number are very narrow.* Consequently, the "end correction," which—strictly speaking—should be applied because the ends are not cut off squarely, will be small. This end correction will in general be too small to cause the vibration curve to deviate appreciably from semicircular form. However, there is a case which requires separate consideration. Suppose the section of the zone is a very small section, constituting only a minute portion of a complete circumference. Then, when the ends are not cut off squarely, the vibration curve may no longer be semicircular. But the vibration curve for a minute section will be minute in extent compared to the vibration curve for the complete zone, and we do not need to concern ourselves over the detail of form of such a minute vibration curve. Consequently, a section of a zone of high number yields *either* a vibration curve which is practically semicircular, *or*, considering a small section of a zone, and then smaller and smaller sections, the vibration curve may eventually come to deviate appreciably

* E.g., the one-thousandth zone will be somewhat less than 1/1000 cm wide when $a=1$ m, $b=1$ m, and $\lambda=5000$ Å as may be shown by eqs. (6) and (7), making use of the fact that $2\pi\rho_n w=\sigma_n$, where w is the width of the zone.

from semicircular form but the curve at the same time becomes so small that it loses importance.

As to the effect produced by a section of a zone of *low* number which terminates in ends which are other than square: A zone of low number has a relatively great width—a width which is of the same order of magnitude as its circumference. Therefore when the ends are other than square, the end correction may cause the vibration curve to deviate considerably from semicircular form, and we can make no general statement regarding the form of the vibration curve. Accordingly, we shall for the present exclude from consideration all cases in which zones of low number are cut off irregularly.

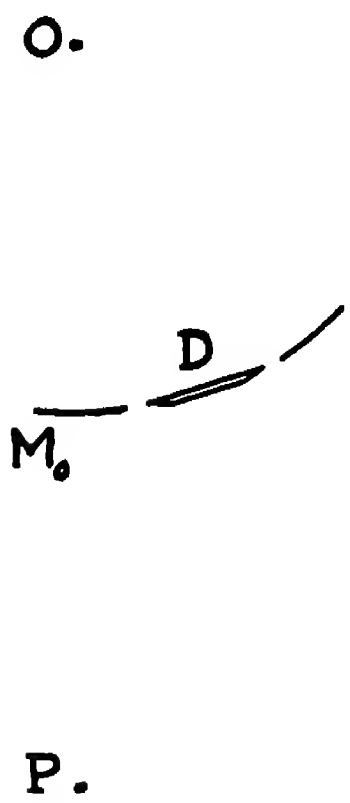


FIG. 35

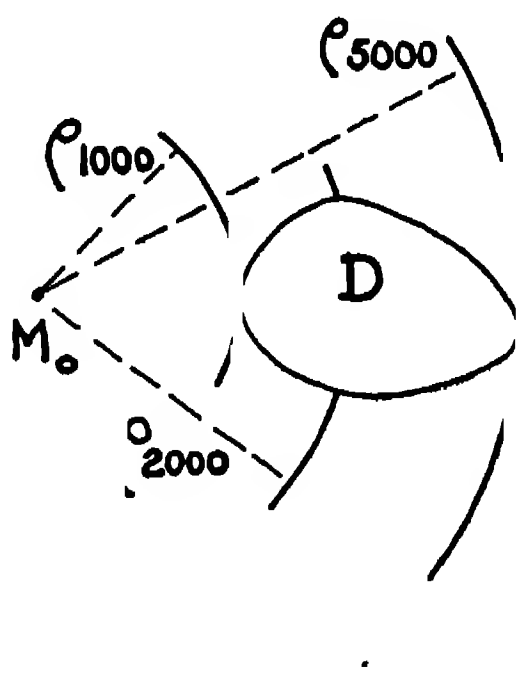


FIG. 36

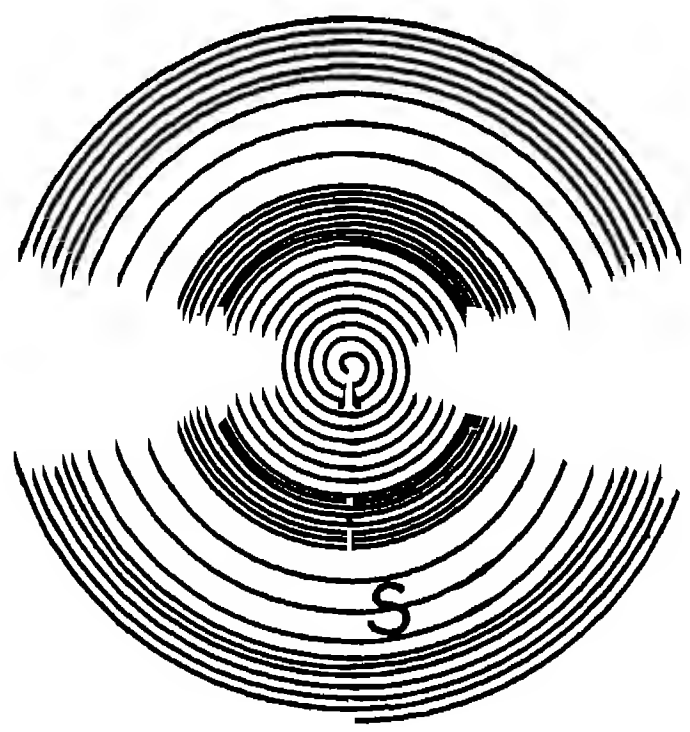


FIG. 37

Referring now to Figure 35, suppose that O is a point source of light, that D is an obstacle, and P is a field point at which we wish to determine the illumination. If no zone of low number is to be cut off irregularly, the field point P must lie well outside of the geometrical shadow of the obstacle. Let us suppose that it does lie well outside. For an observer who takes up a position at P, the pole of the wave will be at M_0 (Fig. 36) and the obstacle will be as represented by D. What form does the complete vibration curve for all of the effective portion of the wave now take and what will be the illumination at P as determined from this vibration curve? Evidently the early convolutions of the curve are unaffected by the presence of the obstacle because no portions of zones of low number are blocked off by it. Suppose that the one-thousandth zone is the first one of which a portion is cut off. This will be a very small portion. Increasing portions of the one-thousand-first, the one-thousand-and-second, etc., zones are blocked off. But large sections of each of these zones are effective. Consequently the vibration curve for each section will be practically semicircular and the only effect will be that

the vibration curve considered as a whole winds up more rapidly than i would if the obstacle were absent (see Fig. 37). The vibration curve con volves throughout in nearly circular arcs about the original center. W finally reach a zone—let us say the two-thousandth—of which a large portion is cut off than of any other. From this point on decreasing portion of successive zones are cut off and the vibration curve winds up *mor slowly* than it normally would, or may even unwind, until the five-thousandth zone is reached, which passes in its entirety. From the five-thousandth zone on the curve winds up in the normal manner. The various stages of winding all take place about the original center and the

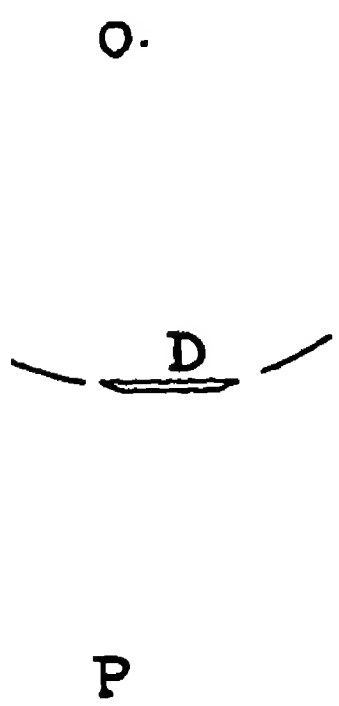


FIG. 38

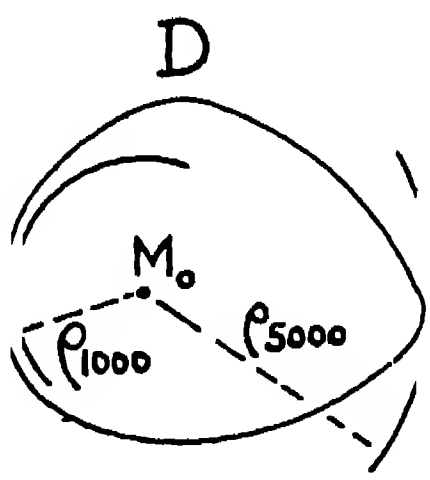


FIG. 39

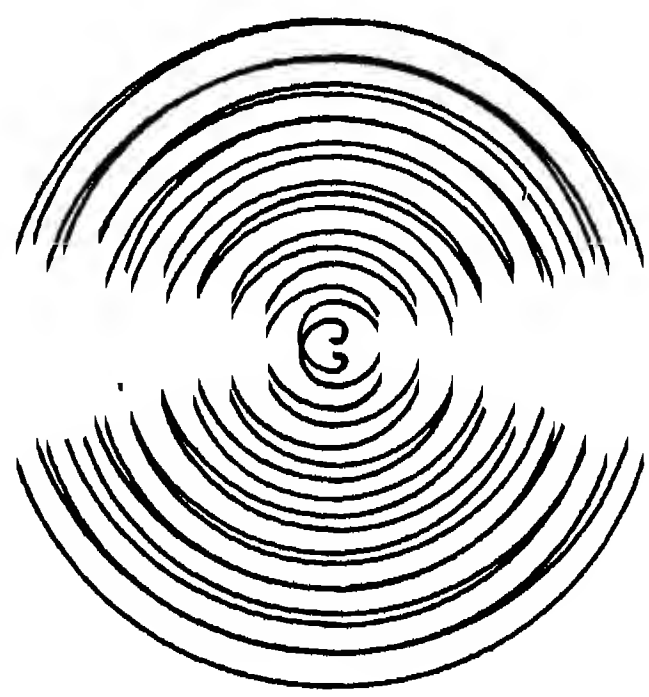


FIG. 40

closing vector S of the curve has the same length as it would have if the obstacle were absent. Consequently all points well removed from the geometrical shadow receive normal illumination.

Let us now consider a field point P (Fig. 38) which is located well within the geometrical shadow of an obstacle D. In Figure 39 the obstacle is represented as it is seen from the field point. The position of the pole of the wave, M_0, is also indicated. Suppose the first thousand zones are blocked off entirely and a minute portion of the one-thousand-and-first and minute but increasing portions of the one-thousand-and-second, the one-thousand-and-third, etc., zones are effective. The minute portions of these zones which are effective will each contribute a minute semiconvolution of undetermined shape to the vibration curve (see Fig. 40). As we pass to zones of higher number, increasing portions are effective. The semiconvolutions of the vibration curve become larger and approach more and more nearly the form of semicircles, unwinding always about a center which must lie very close to

the point where the curve begins. Let us say the five-thousandth zone is the first one which is entirely unobstructed. From this zone on the vibration curve winds up normally, always about the same center, and closes finally at this center which lies very near the beginning of the curve. The closing vector of the entire vibration curve will thus be very minute. We have practical darkness at all points lying well within the geometrical shadow of the obstacle.

We have thus arrived at the law of rectilinear propagation for all points lying well outside of the geometrical shadow and for all points lying well inside of it. The law does not hold for points lying near the edge of the shadow—we shall learn something in later chapters regarding the illumination which obtains near the edge. Finally, the law holds as nearly as it does hold because of the extreme shortness of the wave-length of light. If the wave-length were greater the zones would be larger and it would be necessary to proceed correspondingly farther from the edge of the shadow before the conditions, above outlined, would be such as to yield normal illumination in the illuminated region and practically complete darkness within the shadow.

16. Irregular Obstacle and Circular Aperture.—An interesting prediction can be made from the vibration curve of Figure 40. Suppose an ob-

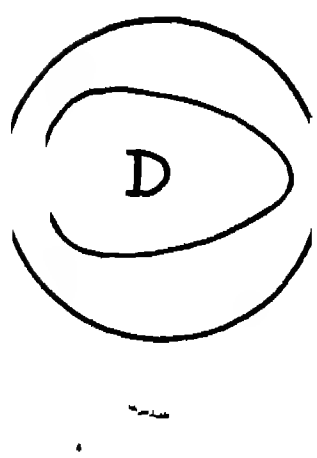


FIG. 41

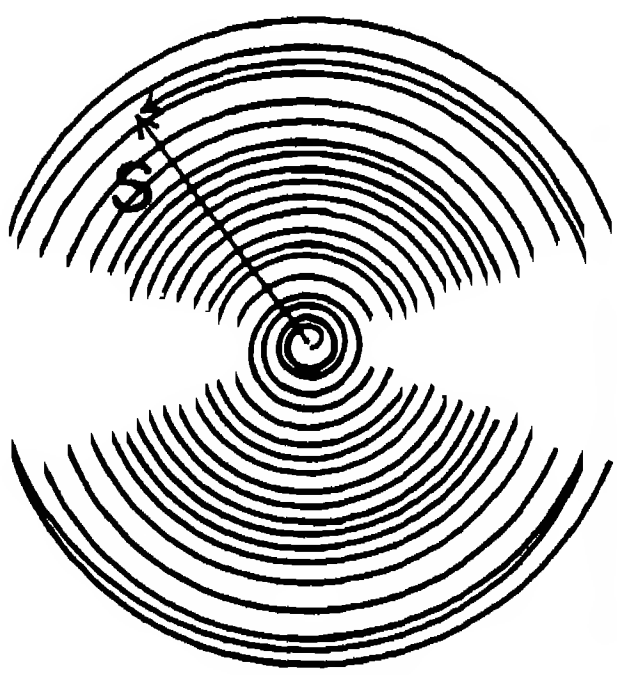


FIG. 42

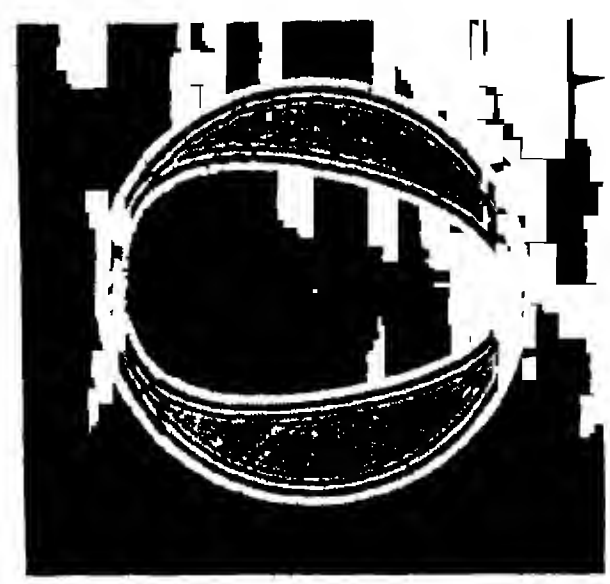


FIG. 43

stacle D (Fig. 41) lies within a circular aperture in an opaque screen. The vibration curve which then obtains for a field point lying on the axis of the aperture is represented in Figure 42. This curve at first unwinds, then it begins to wind up again, and then it terminates at a point determined by the radius of the circular aperture. The closing vector S, drawn from the center outward to the point where the curve ends, represents the illumination at a point on the axis of the aperture, and this illumination will be

practically normal illumination. In other words, in the shadow of an irregular obstacle lying within a circular aperture there should be practically normal illumination on the axis of the aperture. This illumination should obtain irrespectively of whether the aperture alone passes an odd or even number of zones. The prediction is borne out by experiment. Figure 43 shows a photograph in which the illuminated point is readily seen—centered with reference to the aperture.*

* The author is indebted to Mr. George A. Van Lear, Jr., for calling his attention to this prediction and to Mr. Orren Mohler for taking this photograph and several other photographs reproduced in this book as well.

CHAPTER 3

THE METHOD OF LUNES AND THE CORNU SPIRAL

The method which we shall now consider is useful in treating the diffraction pattern due to a single straight edge or that due to a slit or a bar, or, in brief, that due to any object with straight parallel boundaries. Referring to Figure 44, suppose O is a point source of light. This will yield spherical waves. Let the arc $M_3'M_0M_3$ represent the intersection of a wave front with the plane of the drawing. Consider the plane of the drawing as horizontal and let D be a semi-infinite diffracting screen with a straight edge, the edge being supposedly perpendicular to the plane of the paper or in other words vertical. The screen is to be thought of as extending indefinitely, both to the left and in the vertical direction.

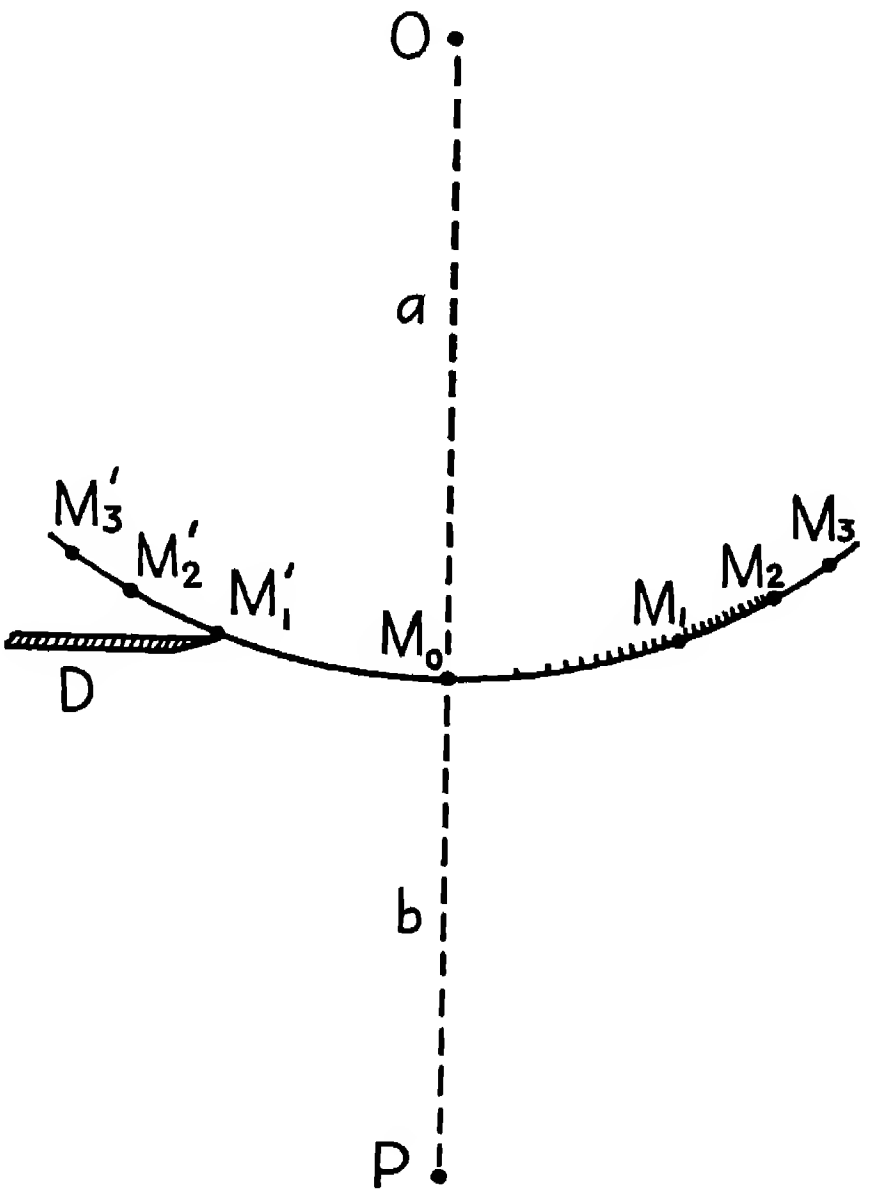


FIG. 44

1. Half-Period Lunes.—Let P be a field point at which we wish to determine the illumination and let M_0 be the pole of the wave with reference to P. Denote the distance OM_0 by a and the distance M_0P by b. Suppose that the light is monochromatic, of wave-length λ, and let M_1, M_2, M_3, etc., be points on the arc such that their distances from P are respectively $M_1P \equiv r_1 = b+\lambda/2$, $M_2P \equiv r_2 = b+\lambda$, $M_3P \equiv r_3 = b+3\lambda/2$, etc. Lay off corresponding points M_1', M_2', etc., on the left of the pole of the wave. At M_0, M_1, M_2 . . . . M_1', M_2', etc., erect vertical great circles on the wave front. These circles will intersect on the perpendicular to the plane of the drawing erected at O. They will delimit so-called "lunes" upon the wave front. The word "lune" is derived from the Latin *luna*, "moon," and signifies a portion of the surface of a sphere which is bounded by intersecting circles. The lunes in question are laid off on the basis of a path increment of $\lambda/2$, and in this respect

they resemble Fresnel zones, which are laid off on the basis of the same path increment. Lunes laid off with $\lambda/2$ as the path increment will be referred to as "half-period" lunes. The width of a given half-period lune is equal to the width of the corresponding Fresnel zone when a, b, and λ are the same in both cases. We have learned that the radii of the Fresnel zones are to each other as the square roots of the natural numbers. Therefore the widths of the zones, and hence also the widths of the lunes, are to each other as the *differences* between the square roots of the natural numbers, that is, as $(\sqrt{1}-\sqrt{0})$ is to $(\sqrt{2}-\sqrt{1})$ is to $(\sqrt{3}-\sqrt{2})$, etc. The widths at first vary rapidly but soon decrease more slowly. The sections of arc M_0M_1, M_1M_2, etc., will be referred to as half-period sections of arc.

2. Elementary Lunes.—Let us begin at the pole of the wave and subdivide each half-period section of arc into a large number m of minute sections so determined that the outer end of each minute section shall be farther removed from the field point P than the inner end by the constant distance $\Delta r = (\lambda/2)/m$. Minute sections so constructed are marked off between the points M_0 and M_2 of our wave front by way of illustration. At each boundary between two minute sections of arc erect a vertical great circle. These circles will outline *elementary* lunes. The first few of these differ markedly from each other in width, but soon the rate of variation becomes less marked.

A given elementary lune contributes a disturbance at the field point P which is proportional in amplitude to the width of the lune. Further, the phase of the resultant disturbance from each elementary lune may be *regarded* as that in which a disturbance originating from the corresponding minute element of arc in the plane of the figure would arrive at P. We shall explain briefly why we may regard the resultant disturbance as having this phase: We are interested in the phase *differences* with which the resultant disturbances from the various elementary lunes arrive at the field point. When we regard the resultant disturbance from each elementary lune as arising from the corresponding minute element of arc, we are omitting from consideration a certain phase angle, but, as will appear in due course, this phase angle is constant for all of the elementary lunes— at least for all of those taken as far from the pole of the wave as we shall ever want to take them. Now the omission throughout of this constant phase angle will not affect the phase *differences*. Therefore we may regard the disturbance from each elementary lune as originating from the corresponding element of the horizontal arc.*

Since we may regard the disturbances as originating from the horizontal

* The question here involved is further discussed in Appendix A, but this Appendix is written on the supposition that the reader will refer to it only after having completed the reading of the present chapter.

arc, our entire problem reduces to one in the two dimensions of the plane of the figure. Let us construct a vibration polygon on the basis of the two-dimensional problem which so arises.

3. Vibration Polygon Based upon Elementary Lunes.—Referring to Figure 45, at the point M_0 of this figure erect a vector, 1, which represents in amplitude and phase the disturbance contributed by the first elementary lune adjacent to the pole of the wave on the right. At the head of this vector erect vector 2, which represents the disturbance from the second elementary lune. Similarly erect vector 3, etc. The elementary lunes which taken collectively constitute the first half-period lune will contribute the first semi-convolution of the vibration polygon extending from M_0 to M_1. The elementary lunes which constitute the second half-period lune will contribute the second semi-convolution of the polygon extending from M_1 to M_2, etc. This polygon will be far from equilateral, especially in the first convolution, but it will be accurately equiangular. The last follows because Δr is constant in constructing the minute elements of arc. As we continue the construction, the vibration polygon will execute smaller and smaller convolutions and will eventually wind up at some point such as J, which will be referred to as the asymptotic point of the polygon or rather of that branch of the polygon which is at present under consideration. A closing vector drawn from M_0 to J would represent the effect of the entire right-hand half of the wave front.

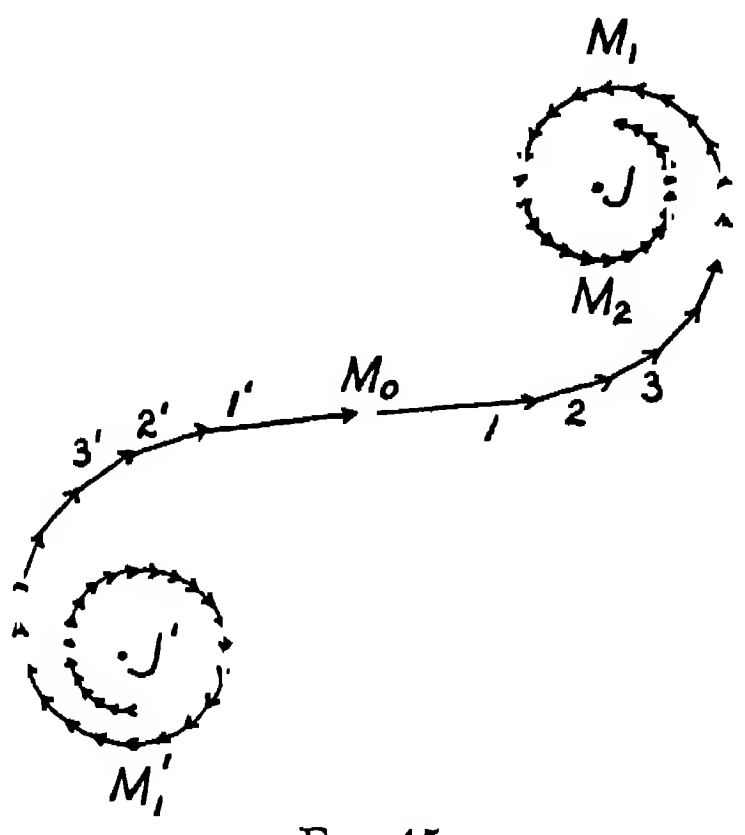


FIG. 45

The minute lunes of the left half of the wave front contribute the vectors 1′, 2′, 3′, etc. It will prove to be convenient in joining these vectors to place the head of each succeeding vector at the foot of the preceding one as illustrated. It is permissible to do this because the order of summation is immaterial. The portion of the polygon extending from M_0 to M_1' represents the contribution from the first half-period lune on the left of the pole of the wave. When a screen with a straight edge is present, with the edge at such a distance from the pole that only the first half-period lune on the left of the pole is effective, but leaving in addition the entire right-hand half of the wave effective, then the resultant disturbance at the field point under consideration will be represented by a closing vector drawn from the point M_1' to the upper asymptotic point J. When, on the other hand, there is no screen present and the entire wave front is effective, the disturbance at the

field point is represented by a closing vector drawn from the lower asymptotic point J' to the upper asymptotic point J. According to the present scheme of dividing the wave front into elementary lunes, the closing vector from J' to J represents the effect of the complete wave front, or what we have called normal illumination.

By considering the number of elementary lunes into which the wave front is divided to increase indefinitely, we may pass from the vibration polygon to a corresponding vibration curve. This vibration curve will prove to be a so-called "Cornu spiral."

We shall next define the Cornu spiral and then proceed to consider first the properties of this spiral from a purely mathematical point of view. The demonstration that the vibration curve is a Cornu spiral will follow later.

4. The Geometry of the Cornu Spiral.—Figure 46 represents the spiral. This is defined by the equation:

$$\psi = \frac{\pi v^2}{2}, \tag{1}$$

where v is the distance measured from the origin of co-ordinates, in terms of the scale indicated along the arc of the curve to a given point on the curve,

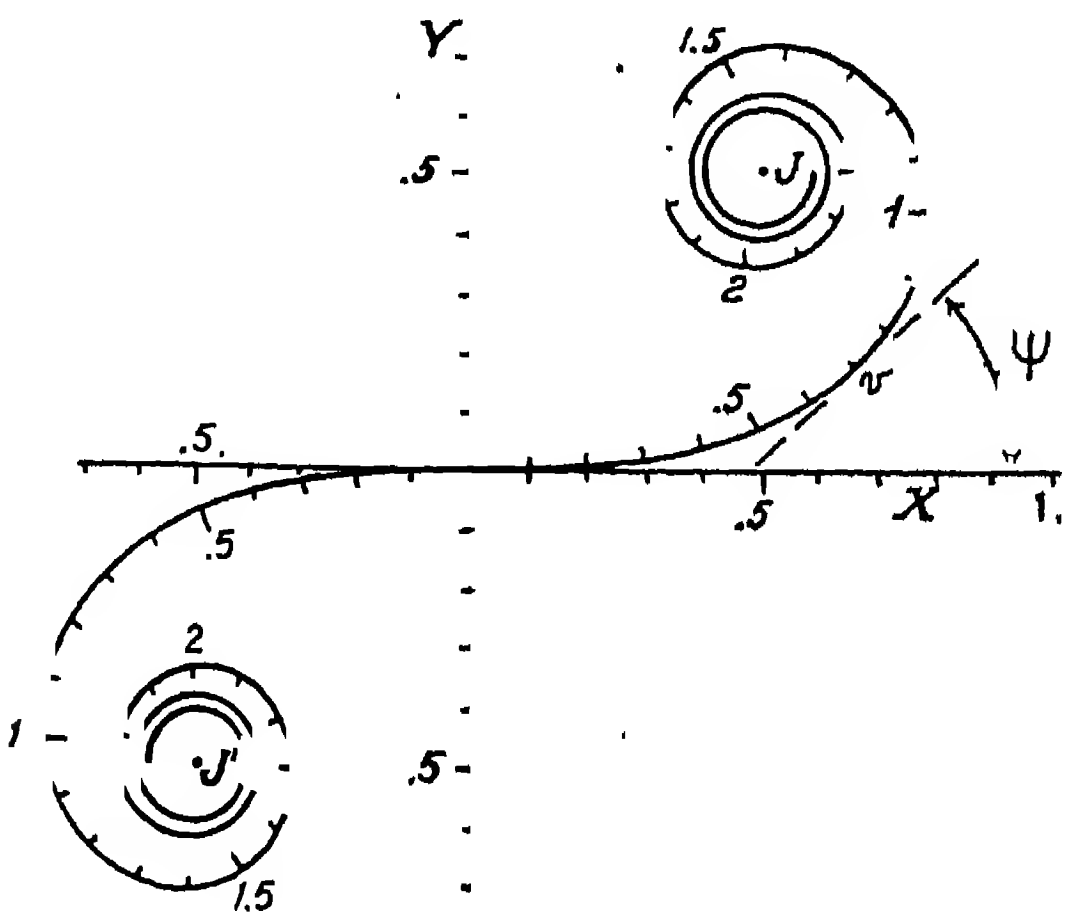


FIG. 46

and ψ is the angle which the tangent erected at this point makes with the X-axis. When $v=0$, the angle $\psi=0$, therefore the curve is tangent to the X-axis at the origin. Values of v measured to the left of the origin are to be considered negative. Since either positive or negative values of v lead to positive values of ψ, there is an inflection at the origin. The curve is vertical

or horizontal, as the case may be, when ψ takes the values $\pi/2$, π, $3\pi/2$, 2π, etc. The corresponding values of v are $\sqrt{1}$, $\sqrt{2}$, $\sqrt{3}$, $\sqrt{4}$, etc. These roots when extracted are 1, 1.41, 1.73, 2, etc., and it is to be noted that the curve in the figure is vertical or horizontal at the points bearing these values according to the scale laid off along the arc. It follows that the lengths of the successive quarter-convolutions of the curve are proportional to the differences between the square roots of the natural numbers, and, as readily deduced, the same relation holds among the successive semiconvolutions and also among the successive complete convolutions. The differences between the square roots of the natural numbers vary rapidly at first, then more and more slowly, and finally approach zero asymptotically. As a result the first convolution of each branch of the spiral is far from circular; but as we proceed to convolutions of higher number, these rapidly become more nearly circular and the spiral winds up about the asymptotic points J and J'.

The curvature of a curve at a given point is defined as the reciprocal of the radius of curvature at that point. This reciprocal is equal to the element of angle swept out by the tangent, as the tangent moves along the curve, divided by the length of the corresponding element of arc. Denoting the radius of curvature at any point v of the spiral by R, the curvature at this point is:

$$\frac{1}{R}=\frac{d\psi}{dv}=\pi v\,, \tag{2}$$

that is, the curvature is directly proportional to the distance v measured along the arc. When this distance is infinite, the curvature is infinite, that is, the radius of curvature is zero.

The intrinsic equation of the curve, equation (1), does not lend itself well to plotting, but the X- and Y-co-ordinates of any desired number of points may be obtained as follows: Suppose that v (Fig. 47) is a point on the spiral, and consider this point to be displaced by an infinitesimal distance dv. The corresponding increments of the X- and Y-co-ordinates of the point are then:

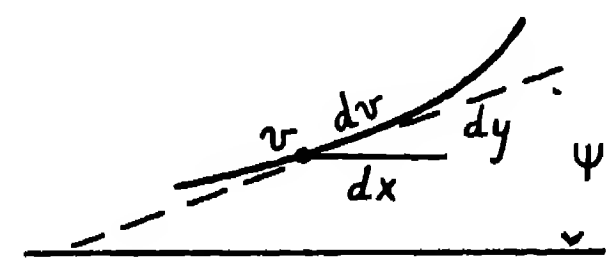


FIG. 47

$$\begin{aligned} dx&=\cos\psi\,dv=\cos\frac{\pi v^2}{2}\,dv\,,\\ dy&=\sin\psi\,dv=\sin\frac{\pi v^2}{2}\,dv\,. \end{aligned} \tag{3}$$

By integrating these equations from zero to an upper limit v_1, we obtain for the co-ordinates of a point lying out along the curve a distance v_1:

$$x_1 = \int_0^{v_1} \cos \frac{\pi v^2}{2}\, dv \text{ and } y_1 = \int_0^{v_1} \sin \frac{\pi v^2}{2}\, dv\,. \tag{4}$$

These integrals are known as the "Fresnel integrals." Fresnel arrived at them in the year 1818 in the development of his theory of diffraction and tabulated their values for various values of the upper limit v_1. The evalua-

TABLE I*

FRESNEL INTEGRALS

v_1	$\int_0^{v_1} \cos \frac{\pi v^2}{2}\, dv$	$\int_0^{v_1} \sin \frac{\pi v^2}{2}\, dv$	v_1	$\int_0^{v_1} \cos \frac{\pi v^2}{2}\, dv$	$\int_0^{v_1} \sin \frac{\pi v^2}{2}\, dv$
0.0	.0000	.0000	2.6	.3389	.5500
0.1	.1000	.0005	2.7	.3926	.4529
0.2	.1999	.0042	2.8	.4675	3915
0.3	.2994	.0141	2 9	.5624	.4102
0.4	.3975	.0334	3.0	6057	4963
0 5	.4923	.0647	3.1	.5616	.5818
0.6	.5811	.1105	3.2	.4663	.5933
0.7	.6597	.1721	3.3	.4057	.5193
0.8	.7230	.2493	3 4	.4385	.4297
0.9	.7648	.3398	3.5	.5326	.4153
1.0	.7799	.4383	3.6	.5880	.4923
1.1	.7638	.5365	3.7	.5419	.5750
1.2	7154	.6234	3.8	.4481	.5656
1.3	6386	.6863	3.9	.4223	.4752
1.4	.5431	.7135	4.0	.4984	.4205
1.5	4453	.6975	4.1	.5737	.4758
1.6	.3655	.6383	4.2	.5417	.5632
1.7	.3238	.5492	4.3	.4494	.5540
1.8	.3337	.4509	4.4	.4383	.4623
1.9	.3945	.3734	4.5	.5258	.4342
2.0	.4883	.3434	4.6	.5672	.5162
2.1	.5814	.3743	4.7	.4914	.5669
2.2	6362	.4556	4.8	.4338	.4968
2 3	.6268	.5525	4.9	.5002	.4351
2.4	5550	.6197	5.0	.5636	.4992
2.5	.4574	.6192	∞	.5000	.5000

* From Gilbert, *Acad. Roy. de Belgique*, **31**, 1, 1863. The values of the cosine integral for $v_1 = 0.1$ and $v_1 = 1.8$ have been corrected following Preston, *The Theory of Light* (5th ed.), p. 296.

tion involves the use of infinite series and is troublesome. It has been carried out by various methods and with greater accuracy since Fresnel's time.† The values are given in Table I. As v_1 increases, each integral passes through

† In succession by Knochenhauer, Cauchy, Quet, Gilbert, and Lommel. The papers of Gilbert, *Acad. Roy. de Belgique*, **31**, 1, 1863, and of Lommel, *Abhand. d. K. Bayer. Akad.*, **15**, 531, 1886, both give detailed references. See also *Brit. Assoc. Report* (Oxford meeting, 1926), p. 273. This *Report* and Lommel's paper give the integrals, not for given values of v_1, but for given values of $\pi v_1^2/2$.

maxima and minima because of the oscillating nature of the trigonometric functions. Moreover, the maxima and minima for the two integrals occur out of step with each other.

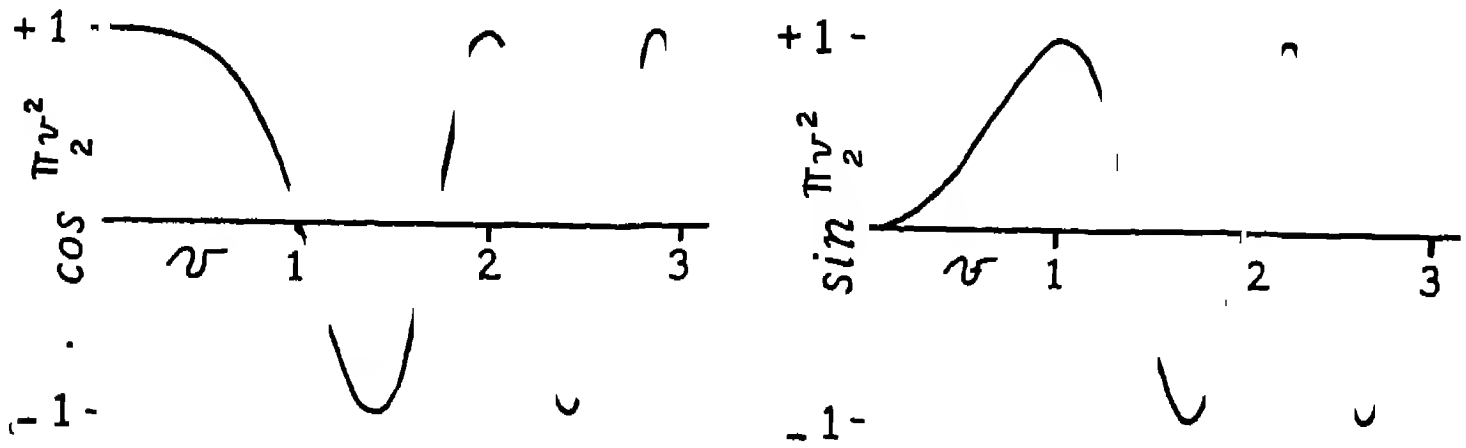


FIG. 48

The integrals may also be evaluated graphically. Referring to Figure 48, suppose v is plotted as abscissa and $\cos \pi v^2/2$ and $\sin \pi v^2/2$ are respectively plotted in separate graphs as ordinates, as indicated. The areas under the curves, taken up to a given value of v, namely, v_1, represent respectively $\int_0^{v_1} \cos(\pi v^2/2)\, dv$ and $\int_0^{v_1} \sin(\pi v^2/2)\, dv$. Areas lying below the axis of abscissae are to be counted as negative. It is to be noted that the points at which the curves cross the horizontal axis crowd closer together as v increases. As the upper limit, v_1, is raised indefinitely the oscillations which the areas, the integrals, undergo, become smaller and smaller. That is, the integrals approach a definite limit as v_1 approaches infinity.

It may be shown analytically* that when the upper limit v_1 is infinite:

$$(5) \qquad \int_0^{\pm\infty} \cos \frac{\pi v^2}{2}\, dv = \pm \tfrac{1}{2} \text{ and } \int_0^{\pm\infty} \sin \frac{\pi v^2}{2}\, dv = \pm \tfrac{1}{2}.$$

The integrals extended to infinity give the co-ordinates of the asymptotic points J and J' of the Cornu spiral which lie respectively at $x = \frac{1}{2}$, $y = \frac{1}{2}$ and $x = -\frac{1}{2}$, $y = -\frac{1}{2}$.

From a practical, a physical point of view, the value of either integral for any very high finite value of v_1, let us say $v_1 = 10{,}000$ may obviously be considered equal to the value for $v_1 = \infty$. Or, speaking in terms of the spiral, the convolutions of very high number are minute in extent and consequently have no physical importance.

The idea of plotting the integrals and of using the graph so obtained for the graphical solution of diffraction problems was first conceived in 1874 by A. Cornu,† and for this reason the spiral bears his name. This innovation

* Drude, *The Theory of Optics*, trans. Mann and Millikan (Longmans, 1917), p. 188, or *ibid.* (3d German ed.), p. 178.

† *Jour. de phys.*, **3**, 1 and 44, 1874.

introduced by Cornu led to the very valuable conception of the graphical treatment of diffraction.

It is to be noted that in the intrinsic equation for the spiral, $\psi = \pi v^2/2$, the value of the constant, $\pi/2$, determines the scale which must be laid off along the curve and along the axes of co-ordinates. When a different constant is used this scale will be different, but the curve will still have the same form and may quite properly be called a "Cornu spiral."

5. The Cornu Spiral as a Vibration Curve.—It was previously stated that the vibration curve based upon the division of the wave front into lunes would prove to be a Cornu spiral. This statement will now be somewhat amplified, and then the needed proof will be given.

When the wave-length is short, as it is in the case of light-waves, the vibration curve proves to be a Cornu spiral from a practical point of view. As will be shown, the early convolutions of the vibration curve approximate extremely closely to the corresponding early convolutions of the Cornu spiral. We do not care whether a high degree of approximation between the two curves exists, convolution for convolution, among the convolutions of high number or not. However, both curves soon convolve in nearly circular arcs and both curves wind up asymptotically about the centers of these nearly circular arcs. Therefore the asymptotic points of the two curves very nearly coincide. We shall make use only of the convolutions of low number and of the asymptotic points—for which high approximation to identity exists. Therefore it is permissible to consider the vibration curve to be a Cornu spiral.

We shall now prove that the early convolutions of the vibration curve approximate extremely closely to the early convolutions of the spiral.

The basic principles of Fresnel's theory of diffraction are expressed by our former equation (13) in chapter 2, which is repeated here:

$$ds = \frac{k A \, d\sigma}{ar} \cos 2\pi \left(\frac{t}{\tau} - \frac{a}{\lambda} - \frac{r}{\lambda} \right). \tag{6}$$

Here $d\sigma$ is a minute element of area on the wave front and r is the distance of this element from the field point under consideration. The element must be one of which all portions may be considered to lie at the same distance r from the field point. The various portions of an elementary lune do not satisfy this condition, hence it would not do at all to let $d\sigma$ represent the area of an elementary lune. However, we may write $d\sigma = h \, dq$, where dq is the width of the elementary lune where it crosses the central horizontal plane, or, what is the same, the length of the corresponding minute element of the horizontal arc drawn across the wave front, and h is a certain "equivalent height."*

* Further discussed in Appendix A.

We will be concerned with only, let us say, the first five, ten, or at most twenty convolutions of the vibration curve. Now the first twenty convolutions will be contributed by the first forty half-period lunes of the wave front, and a simple calculation shows (details given in footnote)* that when $a=1\text{m}$, $b=1\text{m}$, and $\lambda=5000$ Å, these forty half-period lunes will all fall within a horizontally measured distance of only .316 cm from the pole of the wave. Correspondingly, the distance r will vary only from $r=b=100.000$ cm to $r=b+40\ (\lambda/2)=100.001$ cm. That is, r will vary by only one part in one hundred thousand. Now it is only the *proportionate* variation of r which matters in the amplitude term of the foregoing equation. Consequently we shall still obtain a result which is correct to a very high degree of approximation if in the amplitude term of equation (6) we introduce r as a constant equal, namely, to b. This permitted simplification will shortly be introduced.

In the phase term of equation (6) the small variation of r must be taken account of. In this term this small variation is divided by the likewise small quantity λ and is the all-important thing with which we are concerned.

Since the first forty half-period lunes all fall within a small distance on either side of the pole, the variation of the inclination factor over this small range will be extremely slight, and hence we may regard this factor as a constant in developing the theory (illustration given in footnote).†

Where the radius vector r occurs in the phase term of equation (6), let us for convenience substitute for it $b+\delta$, where δ is the retardation, or path increment. Also substituting $d\sigma=h\,dq$ and $r=b$ in the amplitude term as above suggested, equation (6) becomes:

$$ds=\frac{kAh\,dq}{ab}\cos 2\pi\left(\frac{t}{T}-\frac{a+b}{\lambda}-\frac{\delta}{\lambda}\right). \tag{7}$$

Let us now use this equation as a basis for plotting a vibration curve.

Beginning at the pole of the wave, each element dq of the horizontal arc drawn across the wave front goes over into an element dv of the vibration curve, and $dv=C\,dq$, where C is a constant because all terms in the ampli-

* At the outer edge of the fortieth half-period lune the path increment is $\delta=40\lambda/2=.001$ cm, when $\lambda=5000$ Å. The distance measured across the wave front is $q=\sqrt{2ab\delta/(a+b)}$, according to eq. (5) of chap. 2. When $a=1\text{m}$ and $b=1\text{m}$, we have $q=\sqrt{2\times(100)^2\times.001/200}=.316$ cm.

† Taking the same values of a, b, and λ as before, the first forty half-period lunes all fall within .316 cm of the pole. An arc of .316 cm on the wave front subtends an angle of 11 minutes at the source of light. At the end of this arc the inclination angle is $2\times11=22$ minutes. For purpose of illustration let us suppose that k varies in proportion to the cosine of the inclination angle. Accordingly, the variation of k in passing from the pole of the wave to the end of this arc will be only one part in fifty thousand, as may be verified by looking up the cosine function in trigonometric tables.

tude term of equation (7) are constant, except k, and k may be regarded as constant. Moreover, the constant C may be arbitrarily chosen because the scale to which we plot the vibration curve is a matter of choice. Further, let us take $v=0$ when $q=0$, then:

$$v=Cq\,. \tag{8}$$

That is, there is a direct proportionality between the distance q measured across the wave front and the distance v measured along the arc of the vibration curve.

The disturbance from an element of the horizontal arc taken at the pole of the wave, where $\delta=0$, is to be plotted as an infinitesimal horizontal vector. Consequently, the disturbance from an element taken where the path increment δ differs from zero must be plotted, making an angle $2\pi\delta/\lambda$ with the horizontal. That is, the corresponding element of the vibration curve must make an angle $\psi=2\pi\delta/\lambda$ with the horizontal. We may express δ in terms of q, that is, $\delta=q^2(a+b)/2ab$, according to an approximate relation previously deduced (eq. [5] of chap. 2) and then in turn substitute for q the value v/C. Doing this we have:

$$\psi=\frac{2\pi\delta}{\lambda}=\frac{2\pi}{\lambda}\,\frac{q^2\times(a+b)}{2ab}=\frac{\pi v^2}{2}\cdot\frac{2(a+b)}{C^2ab\lambda}\,. \tag{9}$$

That is, ψ is proportional to v^2—or, in other words, we have a Cornu spiral. In order to have on our spiral the scale given by Cornu, we must choose C in such a manner that $2(a+b)/C^2\,a\,b\lambda$ equals unity or:

$$C=\sqrt{\frac{2(a+b)}{ab\lambda}}\,. \tag{10}$$

Substituting this value of C in equation (9), we have:

$$\psi=\frac{2\pi\delta}{\lambda}=\frac{\pi v^2}{2}\,. \tag{11}$$

This is identical with the defining equation for the Cornu spiral.

When the constant of proportionality connecting q and v, namely, C, is chosen as above, it is possible to apply the spiral, with the scale given by Cornu, to any diffraction problem of suitable nature, irrespective of particular values of a, b, and λ.

Since $v=Cq$, it follows from equation (10) that:

$$v=q\sqrt{\frac{2(a+b)}{ab\lambda}}\,. \tag{12}$$

We have now deduced that equation (7), used as a basis for plotting a vibration curve, leads to a Cornu spiral. However, this equation is itself an approximation and there are further approximations introduced in deducing that the equation leads to a Cornu spiral. But these approximations are all extremely good while we are close to the pole of the wave, or in other words for the early convolutions of the vibration curve. It is because the wave-length of light is very short that the approximations are so extremely good for the early convolutions. And for the same reason we are able to reduce what is in its general aspects a very difficult problem in three dimensions to a much simpler problem in the two dimensions of the horizontal plane.

The vibration curve is from a practical point of view a Cornu spiral and the two curves will henceforth be considered as identical.

Suppose that only a given portion of the wave front is effective. To this portion will correspond a portion of the spiral extending from a certain point v_1 (Fig. 49) to a second point v_2. Passing over for the time being the

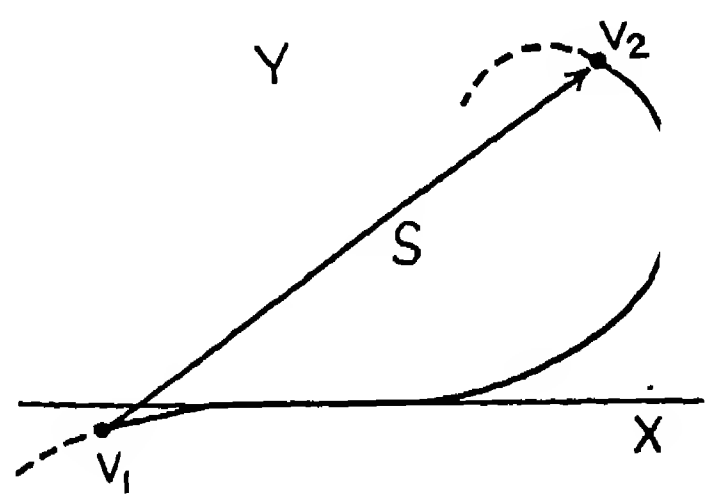


FIG. 49

question of how the positions of the points v_1 and v_2 may be determined, be it noted that the closing vector S from v_1 to v_2 represents the amplitude of illumination which prevails, and the square of the length S represents the intensity. The vector S may be resolved into a horizontal component $X = \int_{v_1}^{v_2} \cos(\pi v^2/2)\,dv$ and a vertical component $Y = \int_{v_1}^{v_2} \sin(\pi v^2/2)\,dv$. The sum of the squares of these two integrals will now represent the intensity. Thus the intensity of illumination is represented by:

$$I = \left[\int_{v_1}^{v_2} \cos\frac{\pi v^2}{2}\,dv\right]^2 + \left[\int_{v_1}^{v_2} \sin\frac{\pi v^2}{2}\,dv\right]^2. \tag{13}$$

The intensity of illumination is represented by the sum of the squares of the Fresnel integrals taken between appropriate limits.

When the entire wave front is effective, the closing vector becomes the one joining the asymptotic points. The Fresnel integrals are then to be extended from $-\infty$ to $+\infty$.

The reader who is making his first study of the subject of diffraction no doubt feels troubled by many points of the difficult theory over which he has been laboring. However, he is counseled not to try to dispel his difficulties by further applying himself intensively to the present chapter forthwith. He should rather pass on soon to the study of how the Cornu spiral is used in arriving at typical diffraction patterns, as taken up in the next chapter. There many matters will clarify themselves.

Amplified treatment of several points which have arisen in the present chapter is presented in Appendix A. The reader may do well, however, to defer the reading of this Appendix until after having learned how the spiral is used.

CHAPTER 4

TYPICAL DIFFRACTION PATTERNS OF THE FRESNEL CLASS

Fresnel studied the diffraction patterns which are formed when light from a secondary source O (Fig. 50) passes a diffracting screen of some type D, and the pattern is formed in the field plane PP'. This arrangement includes all situations which we have thus far considered. Fresnel used a condensing lens for concentrating a beam of sunlight upon the pinhole or slit which acts as a secondary source, and an eyepiece for viewing the pattern already formed (as described at the beginning of chap. 1). These lenses serve simply as accessories—they play no rôle in forming the pattern. But Fresnel did not place any lenses between the secondary source O and the field plane PP'. Lenses so placed affect the formation of the pattern. However, today we also include as Fresnel patterns those patterns which are formed when one or more lenses are placed between the secondary source and the field plane, *provided* the secondary source and the field plane do *not* occupy *conjugate focal planes*. When they *do* occupy conjugate focal planes the pattern is of another, the so-called Fraunhofer class. The reason for this classification we shall learn in the next chapter after we have studied some patterns of both the Fresnel and the Fraunhofer class. In the meantime we may think of the Fresnel class simply as comprising those patterns which are formed without using lenses between the secondary source and the field plane where they would affect the formation of the pattern.

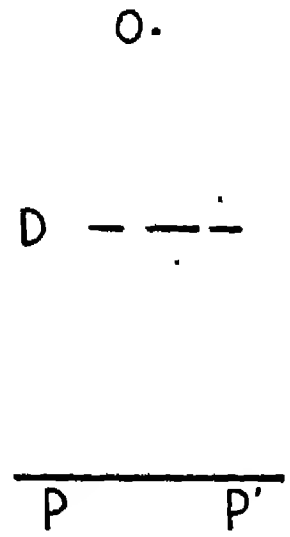


FIG. 50

Patterns formed without lenses were first observed by Grimaldi, yet they have come to be known as Fresnel patterns. Fresnel greatly advanced our knowledge of them, both on the experimental and on the theoretical side. Moreover, Fresnel's work preceded that of Fraunhofer on the other type of pattern by only a few years. The advances made by the two distinguished physicists appear in retrospect as contemporaneous and the designations which we apply to the two classes of patterns co-ordinate these advances.

In undertaking our consideration of Fresnel patterns it would seem reasonable upon first thought to begin with, let us say, the pattern formed by a circular aperture. The method of zones leads, as we have seen, to an elegant and simple determination of the illumination which obtains on the axis of the aperture, and in any field plane perpendicular to this axis the illumina-

tion must be symmetrical about the axial point. However, even though this illumination be symmetrical, it is a difficult matter to deduce theoretically what it should be. The circular disk presents an equally difficult problem. We shall therefore begin by using the Cornu spiral as a vibration curve for the study of the patterns which result from diffracting systems which have straight parallel boundaries. With such systems, when the boundaries are vertical and a point source of light is used, the fringes are vertical and practically straight. Hence it is possible, and it is customary, to use a narrow vertical slit as source. The problem may be treated, as we have learned, as one in a horizontal plane. We shall determine the illumination along a horizontal line, e.g., the line PP' of Figure 50, and it will be understood in each example that the pattern consists of fringes running perpendicular to the plane of the figure. We shall consider first the pattern formed by a single straight edge.

1. The Straight Edge.—The pattern due to a vertical straight edge consists of a single set of vertical fringes (see Fig. 51, upper portion). The diffracting screen extends indefinitely to the left. The fringes occur just beyond the edge of the geometrical shadow, indicated by the point E. They diminish rapidly in spacing and in degree of intensity variation, merging soon into uniform, normal illumination. The lower portion of Figure 51 is a plot of position in the pattern as abscissa against intensity as ordinate. In Figure 52 the same intensity variation is shown to larger scale, curve I, and the variation of amplitude of illumination is shown by curve A.

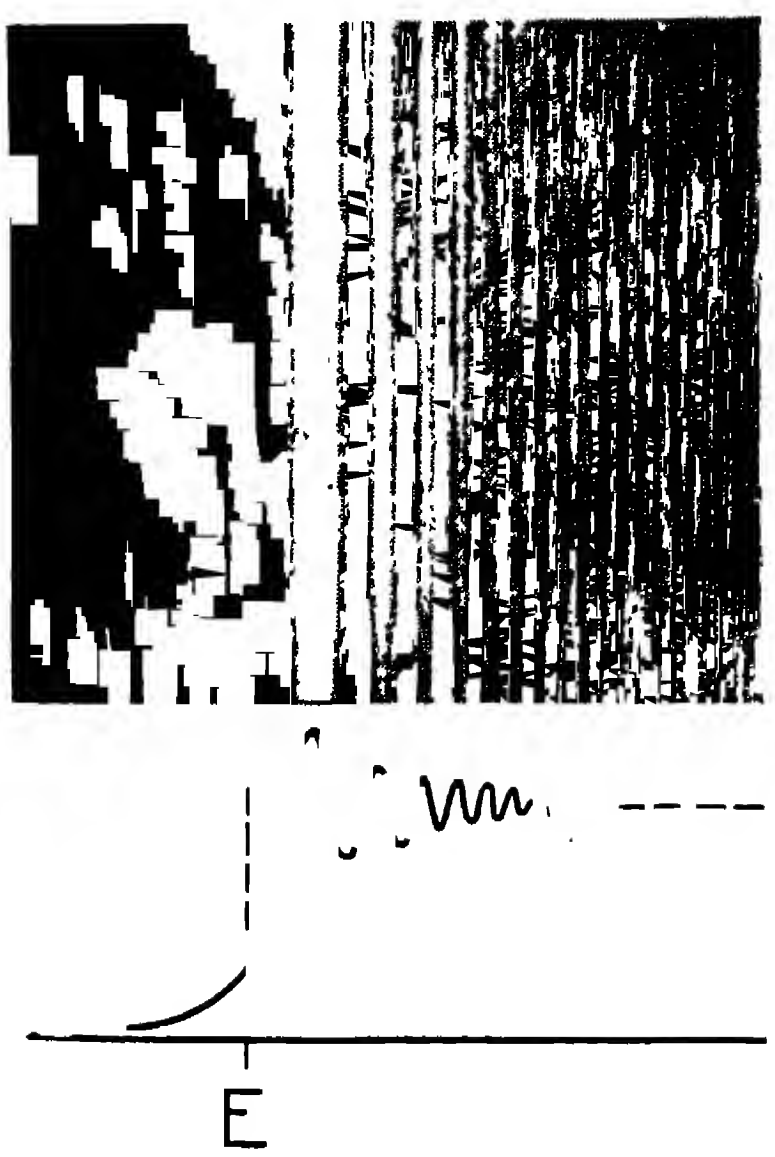


FIG. 51

Our problem consists in accounting for the diffraction pattern by using the Cornu spiral as a vibration curve. Referring to Figure 53, suppose P_0 is a point lying on the edge of the geometrical shadow. The pole of the wave with reference to this point is M_0, coinciding in position with the edge of the screen. The left half of the wave is obstructed and the entire right half is effective. The amplitude of illumination is represented by the vector OJ (Figure 54a) drawn from the origin of the spiral to the upper asymptotic point. This amplitude is half of the of normal illumination and hence, as is to be noted, the curve A (Fig

crosses the edge of the geometrical shadow at one-half of the height which represents the amplitude of normal illumination. Correspondingly, the curve I crosses at one-quarter of the height of normal intensity.

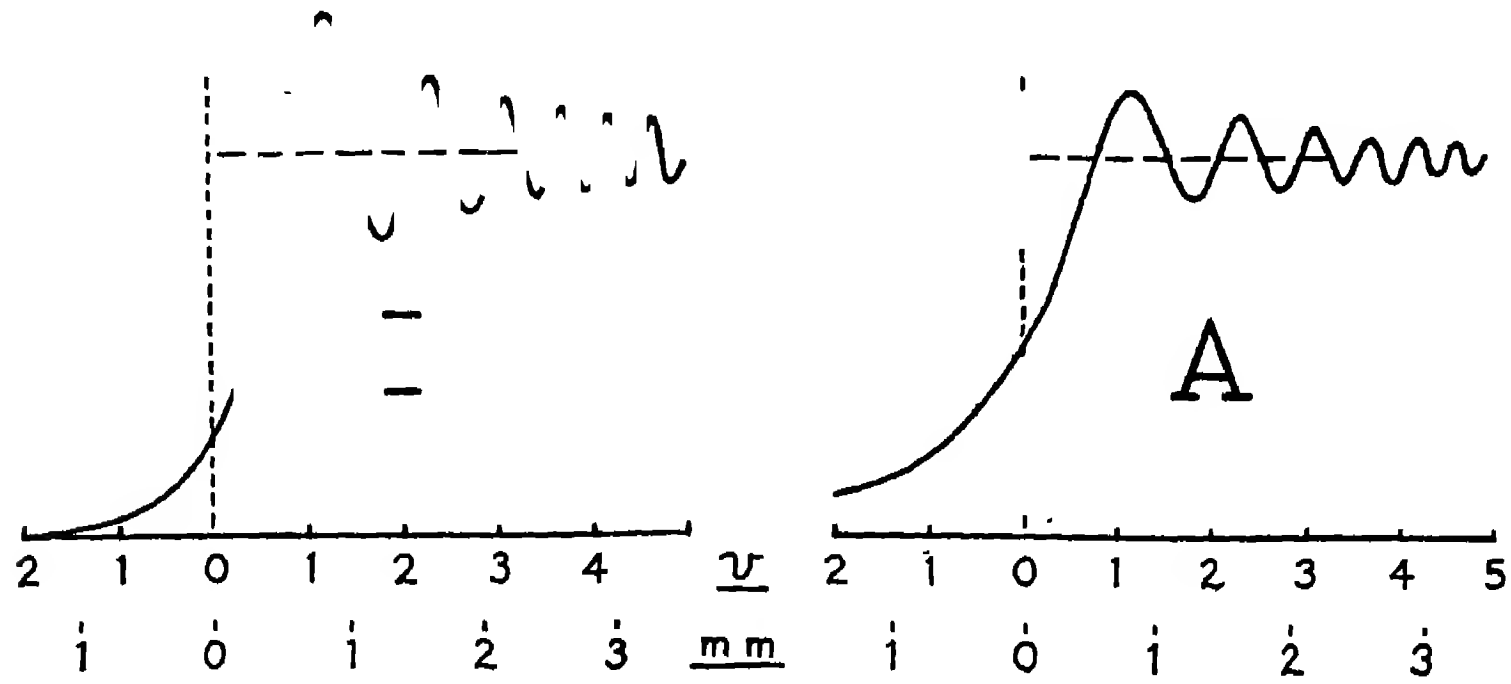


FIG. 52

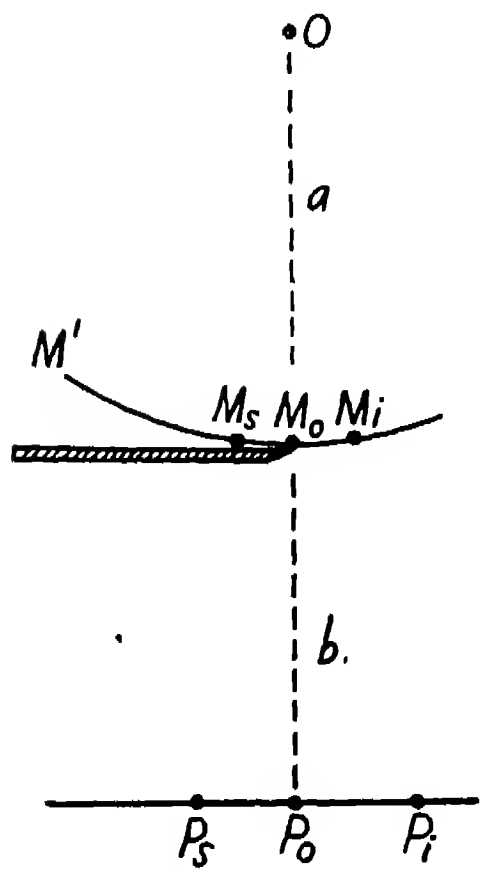


FIG. 53

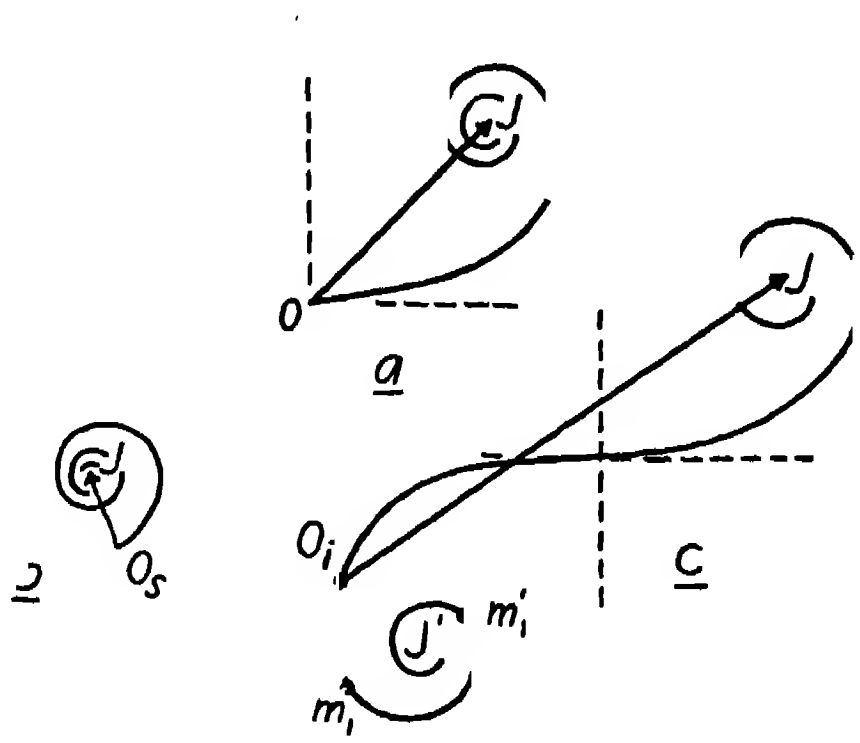


FIG. 54

For the field point P_s (Fig. 53), lying within the shadow, the pole of the wave is M_s. Viewed from P_s, the left half of the wave is M_sM' and this is entirely obstructed, and so is a portion of the right half, M_sM_0, but the remainder of the right half of the wave is effective. The origin of the spiral always corresponds to the pole of the wave which is now M_s. The illumination from the effective portion of the wave is measured by the closing vector O_sJ of the corresponding portion of the spiral, namely, by a vector such as O_sJ (Fig. 54b). If the point P_s were moved to the left, farther into the shadow, the point O_s would move along the spiral toward J, the closing vector be-

coming gradually and continuously shorter. The illumination diminishes *continuously* until complete darkness is attained—there are no fringes within the geometrical shadow—there are no so-called *internal* fringes.

For a point P_i (Fig. 53), lying in the illuminated region of the pattern, the pole of the wave is M_i. Viewed from P_i, a portion of the left half of the wave, namely, M_iM_0, is effective, and so is all of the right half. The illumination is therefore measured by a vector such as O_iJ (Fig. 54c). If the point P_i were moved to the right, the point O_i would move farther out on the left branch of the spiral. When O_i reaches the point m_1, the closing vector, drawn always to the point J, attains a maximum value. This is the first maximum attained and to it corresponds the first bright fringe in the pattern. When O_i reaches the point m_1', the closing vector attains its first minimum value, representing the first dark fringe. As the moving point proceeds farther out along the left branch of the spiral, the closing vector passes through successive maxima and minima representing the successive bright and dark fringes of the pattern, which consists of a single set of so-called *external* fringes. Upon proceeding farther and farther into the illuminated region, the convolutions of the spiral become smaller—the maxima and minima occur closer together and become less and less pronounced. Normal illumination, represented by a vector drawn from J' to J, is gradually attained.

The first maximum in the pattern occurs before the first half-period element of the wave front is fully exposed, as indicated by the fact that the point m_1 (Fig. 54c) occupies a position *on*, and not quite at the end of the first semiconvolution of the spiral. A corresponding conclusion applies to the first minimum and to subsequent maxima and minima.

At a given point in the diffraction pattern the proper point on the spiral must be chosen from which to draw the closing vector to the upper asymptotic point J. It remains to show how the point on the spiral is determined: Referring to Figure 53, denote by the symbol z, the distance of a point in the pattern, e.g., P_i, from the edge of the geometrical shadow, thus $z=P_0P_i$. The pole of the wave with reference to P_i is M_i, and this lies at a distance $q=M_0M_i$ from the straight edge. Since both the distances q and z are small:

$$q=z\frac{a}{a+b}. \tag{1}$$

We have previously derived an equation relating a distance q measured across the wave front with a length of arc of the spiral v, namely, equation (12) of chapter 3, which we repeat here:

$$v=q\sqrt{\frac{2(a+b)}{ab\lambda}}. \tag{2}$$

Eliminating q from equations (1) and (2):

$$v = z\sqrt{\frac{2a}{(a+b)b\lambda}}. \tag{3}$$

To a given displacement z in the diffraction pattern there corresponds a displacement q of the pole of the wave, and an addition or subtraction of a length of arc v to the effective portion of the spiral. The quantities z, q, and v are interrelated according to equations (1), (2), and (3) by factors which are constant once the values of a, b, and λ have been chosen. In Figure 52 the upper of the two scales of abscissae are in values of v. This scale is independent of particular values of a, b, and λ. The lower scales are in values of z in millimeters, plotted on the supposition that $a = 1$ m, $b = 1$ m, and $\lambda = 5000$ Å. It is to be noted that v is expressed in equations (2) and (3) as a ratio of lengths and v hence has the dimension of a pure number.

The verification of the theoretically expected positions of the fringes in the straight-edge pattern was made by Fresnel. This verification today forms an instructive laboratory exercise for which directions are given in Appendix B. In connection with this exercise it is desirable to use a graph of the Cornu spiral plotted to a large scale.* The verification of the intensities of the fringes is a more difficult problem, involving careful photometry. This was carried out recently for the first time by a photographic method.†

As stated in chapter 1, section 2, if we follow the locus of a given fringe, in space, from the diffracting edge backward, the locus is a hyperbola. This fact follows immediately from equation (3). For a given fringe v is constant. For example, for the first bright fringe v has a value of about 1.2. We now regard b as independent variable and z as dependent variable, that is, we regard z as a function of b. We then obtain a hyperbola in b and z.

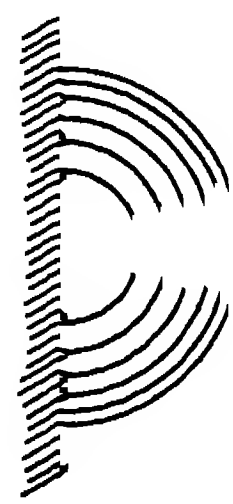


FIG. 55

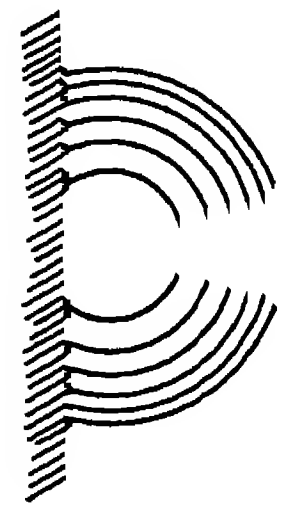


FIG. 56

The occurrence of the fringes in the straight-edge pattern can also be explained qualitatively in a very elementary manner by invoking only the method of zones: For a field point on the edge of the geometrical shadow, the straight edge allows one-half of each zone to pass (see Fig. 55). The amplitude of illumination

* This may be plotted from the table of Fresnel integrals given in chap. 3, or copies may be obtained for a small sum from the Blue Print Shop, Engineering College, University of Mich., Ann Arbor, Mich.

† T. N. Lyman, *Proc. Nat. Acad.*, **16**, 71, 1930.

is hence half that of normal illumination and the phase of the disturbance coincides with that due to the first zone. Upon moving out from the edge of the geometrical shadow a small distance, a larger proportion of the first zone is added than is added for zones of higher number (see Fig. 56). Consequently the illumination is increased and attains a maximum under the conditions represented in the figure. Upon moving farther out, relatively more of the second zone is added, so that the illumination decreases. In this manner the occurrence of the successive maxima and minima may be explained. It is difficult, however, to formulate a corresponding, simple explanation by the method of zones for the continual decrease in illumination which occurs upon proceeding inward from the edge of the geometrical shadow.

Through accounting for the pattern due to the straight edge, whether by the Cornu spiral or otherwise, we have incidentally again deduced the law of the rectilinear propagation of light, to the degree to which this law holds, for we find that complete darkness should obtain well within the geometrical shadow and normal illumination should obtain well outside of the shadow. Similarly in the subsequent consideration of patterns arising from various other diffracting systems the reader should bear in mind that whenever the nature of a pattern is deduced from theory, this deduction in each case carries with it a deduction of the law of rectilinear propagation to the degree to which the law holds.

External fringes similar in general appearance to those formed by a straight edge surround the shadow of almost any object when a point source of light is used. This is more or less what we should expect, because general considerations of physical continuity lead us to anticipate that if an edge deviate slightly from being straight the fringes will curve correspondingly but will in other respects retain their general appearance. The question then arises: How much may an edge be curved before the general appearance changes? The answer given both by observation and by theory, in those cases which lend themselves to analysis, is that an edge may permissibly have fairly great curvature, and that external fringes of the general type of those formed by a straight edge will ordinarily surround all portions of the shadow of an object except where the curvature of the edge is very great, as, for example, where the object has a sharp projection. However, very small objects fail to show the characteristic external fringes in question, for obvious reasons, and the same applies to objects which are very small in one cross-sectional dimension, as, for example, a fine wire or a hair.

2. The Single Slit.—The nature of the pattern which arises from a slit depends markedly upon the width of the slit. Examples are shown diagrammatically in Figure 57. The slit width increases in the successive dia-

grams, progressing downward. The edges of the geometrical shadow and the amplitude of normal illumination are indicated by broken lines. It is to be noted that at the center of the pattern the illumination is in some cases a maximum and is in some cases a minimum.

Let us see how the Cornu spiral is to be used as a vibration curve:

Referring to Figure 58, at the central field point P_0, equal portions of the wave front to the right and left of the pole are effective. Corresponding equal arcs of the spiral must be taken along each branch (Fig. 59*a*). The length of the closing vector, drawn from O_1 to O_2, is a measure of the illumination which obtains at the central point of the pattern. The length of arc of the spiral which is to be taken will be denoted by $2v'$. Thus v' denotes the length along each branch. The value of v' is determined by equation (2), in which q is to be given the value $q' = w/2$, where w is the width of the slit.

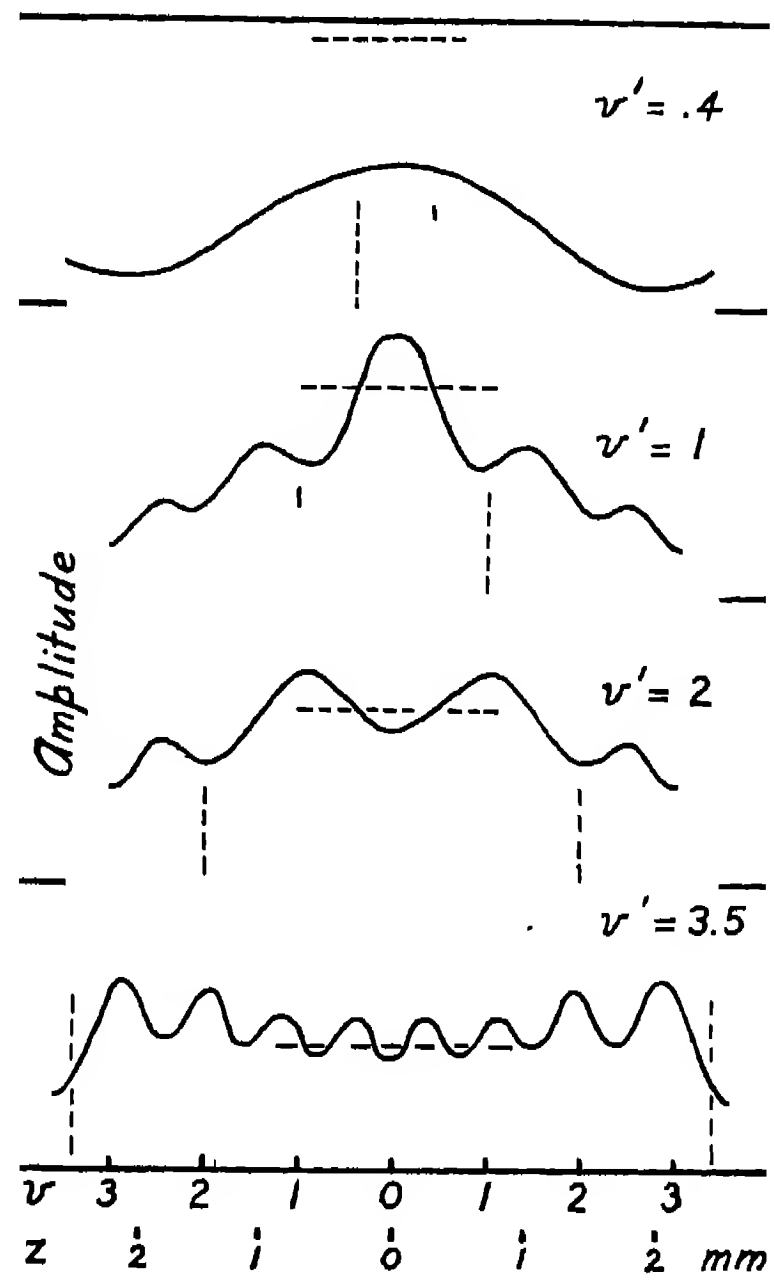


FIG. 57

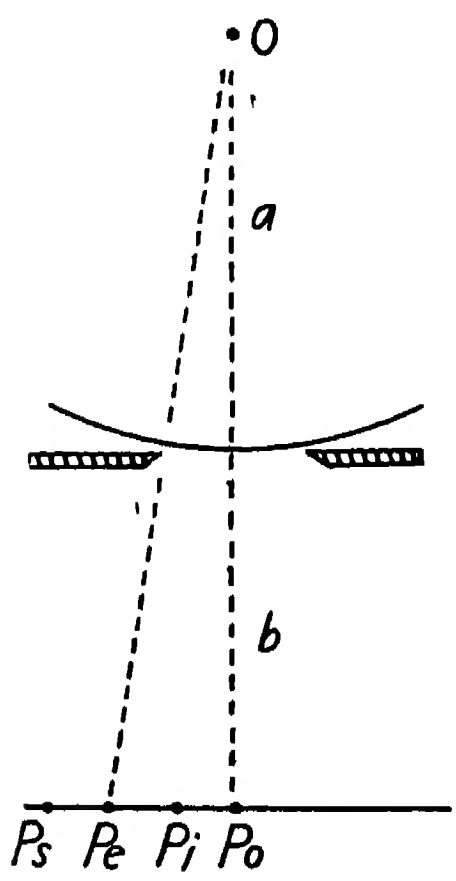


FIG. 58

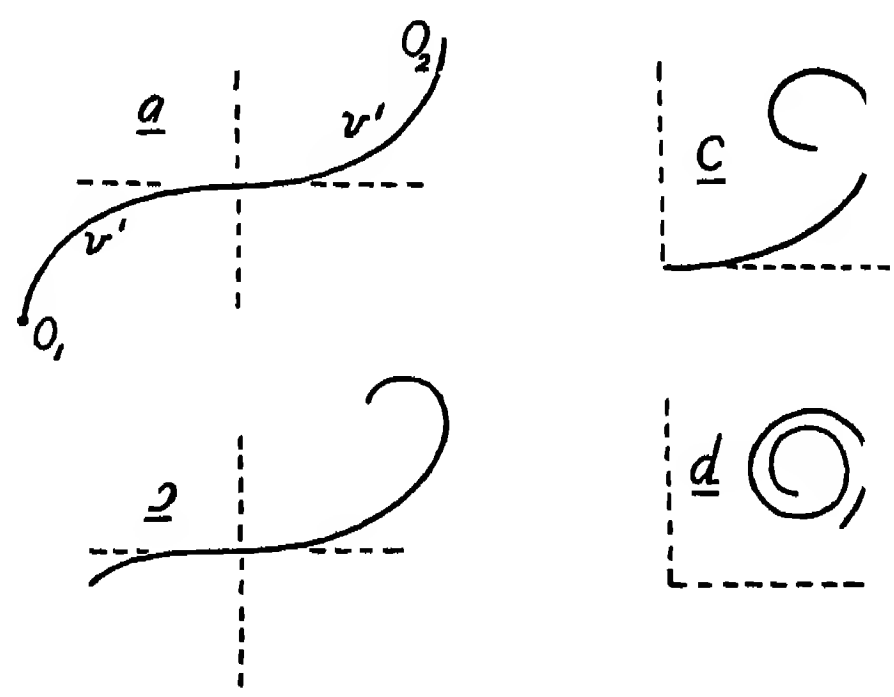


FIG. 59

To determine the illumination which obtains at another field point, e.g., P_i (Fig. 58), the same total length of arc of the spiral $2v'$ must again be used because the same width of the wave front, determined by the width of the slit, is still effective. The pole of the wave has now shifted to the left, however—a smaller part of the effective portion of the wave lies on the left of the pole and a larger part lies on the right. Consequently the length of arc $2v'$ must now extend a lesser distance along the left branch of the spiral and a correspondingly greater distance along the right branch (see Fig. 59*b*). The closing vector of this arc represents the amplitude of illumination. The amount by which the arc is shifted along the spiral is determined by equation (3), in which the symbol z represents the distance of the field point in question from the central point of the field.

When the given field point lies on the edge of the geometrical shadow on the left, that is, at P_e (Fig. 58), the effective portion of the wave front extends from the pole toward the right. The length of arc $2v'$ must now extend from the origin of the spiral along the right-hand branch (Fig. 59*c*). When the field point lies within the geometrical shadow on the left, at P_s, the effective portion of the wave front begins at a certain point on the right of the pole and extends farther to the right—the length of arc $2v'$ now extends from a certain point on the right-hand branch of the spiral (Fig. 59*d*) to another point lying farther out on this branch. The closing vector always represents the amplitude of illumination. The curves of Figure 57 were plotted from point to point in the manner which has just been indicated. The scales of linear distances, z-values, are plotted on the supposition that $a=1$ m, $b=1$ m, and $\lambda=5000$ A.

When the width of the slit is small, topmost curve of Figure 57, the corresponding segment of the spiral of length $2v'$, is short, and it is necessary to shift this segment far out along one branch or the other of the spiral before it curls so much that the closing vector becomes markedly shortened. In other words, the central maximum of the pattern extends far beyond the confines of the geometrical beam, and the narrower the slit is, the farther does the central maximum extend. We approach the ideal condition of an infinitely narrow portion of the wave front giving rise to Huygens' wavelets which spread out laterally in all directions.

When the width of the slit is intermediate, there are fringes both within the geometrical shadow, internal fringes, and fringes outside of this, external ones. The division of the fringes into two sets is, however, a rather arbitrary one as there is no abrupt transition in spacing or intensity as we pass across the edge of the geometrical shadow. In contradistinction we shall find that for the case of a *bar* of intermediate width, a rather abrupt transition occurs.

When the width of the slit is great (see lowermost curve of Fig. 57) the

corresponding segment of the spiral of length $2v'$ is long, and comprises numerous convolutions in each branch of the spiral when the central field point is under consideration. Upon leaving the central field point, only small variations from normal illumination occur at the outset. As either edge of the geometrical shadow is approached, the variations increase in magnitude. As the slit is progressively widened farther, the pattern progressively approaches that of two opposed straight edges acting independently of each other.

3. The Bar.—The problem of the bar forms a complement to the problem of the slit. The entire wave front is now effective except the portion which is stopped by the bar. At the central point of the shadow, equal portions of the wave front to the right and left of the pole are stopped. Corresponding equal arcs must be *removed* from each branch of the spiral,

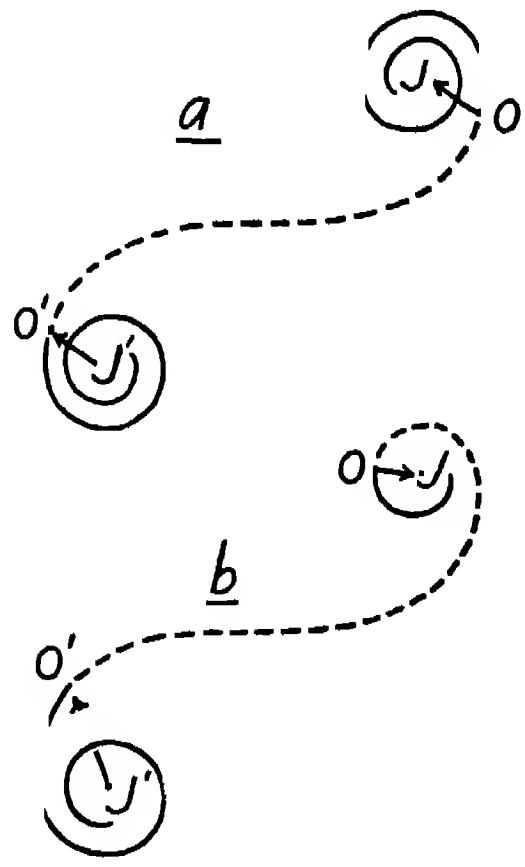


FIG. 60

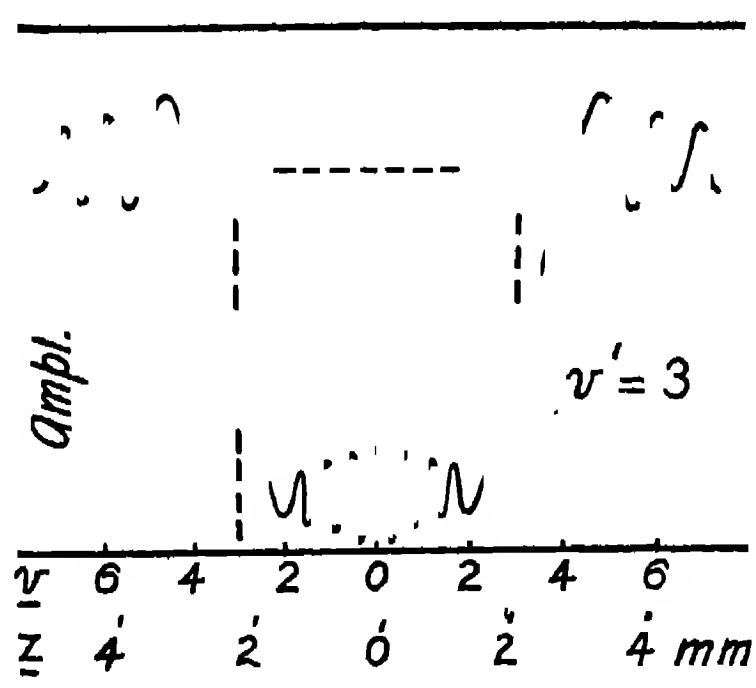


FIG. 61

namely, the broken portion of the curve in Figure 60*a*. The illumination is obtained by adding the vectors OJ and $J'O'$. The length of the broken arc will be denoted by $2v'$, so that v' denotes the length along each branch.

To determine the illumination which obtains at a field point other than the central one behind the bar, the same total length of arc of the spiral, $2v'$, must again be removed (Fig. 60*b*), but since the pole of the wave does not now lie at the center of the bar, the arc $2v'$ is now no longer located symmetrically with reference to the origin of the spiral. The vectors OJ and $J'O'$ differ in length and direction and they must be added *vectorially* to obtain the amplitude of illumination at the field point in question. By applying equations (1), (2), and (3) in accord with the principles which were outlined in discussing the cases of the straight edge and the slit, the complete diffraction pattern may be plotted from point to point.

The length of arc which has been removed in drawing Figures 60*a* and 60*b* corresponds to a bar which may be termed a narrow one from the point of view of the diffraction pattern which results. For a bar of medium width several convolutions of the spiral would have to be removed. Figure 61 represents the pattern for such a bar. The length of arc of the spiral which was removed in arriving at this pattern is $2v' = 6$, or two and one-quarter convolutions of each branch, at the central point of the pattern.*

There is a set of external fringes on each side of the shadow and each set is nearly identical with the set of fringes which would be produced by a single straight edge when the values of a, b, and λ are the same in both cases. That this should be true may be readily seen. Consider the field point which lies on the right-hand edge of the geometrical shadow. The vector which corresponds to OJ in Figure 60 now extends from the origin of the spiral to the upper asymptotic point. The removed arc of the spiral extends from the origin along the lower branch for six units measured along the arc. The vector which corresponds to $J'O'$ is then very short. Upon proceeding to the right in the pattern into the illuminated region, the removed arc of the spiral slides farther out along the the lower branch. The vector which corresponds to OJ undergoes the same variations which the vector representing the external illumination at a single straight edge undergoes, and the vector which corresponds to $J'O'$ is so short that it contributes but little to the final result.

There is also a set of internal fringes. The central fringe of this set is always a bright one irrespective of the width of the bar, because the vectors OJ and $J'O'$ always point in the same direction at the central position, and a displacement from the central position causes these vectors to rotate in opposite directions with an ensuing diminution of their sum which measures the illumination. In contradistinction, at the center of the fringe system due to a slit, the illumination may be either a maximum or a minimum. It will presently be shown that the spacing of the internal fringes in the diffraction pattern due to a bar should not depend upon the distance which we have designated by a, the distance from the source to the bar, but should be determined solely by the magnitudes of b, λ, and w, the width of the bar—and this is borne out by experiment. In fact, the internal fringes have the same spacing which the interference fringes in Young's experiment with two slits have when the slits are separated by a distance equal to the width of the bar. The fringes may be regarded as arising from the interference of light inflected at the two edges of the bar. This viewpoint will be further discussed in chapter 7, sections 10 and 11.

* Taking $a = 1$ m, $b = 1$ m, and $\lambda = 5000$ Å and substituting the value $v' = 3$ for v in eq. (2), we find for q the value $q' = .106$ cm, or for the diameter of the bar, $w = 2q' = .212$ cm, and for the width of the geometrical shadow, .414 cm.

When the bar is a narrow one, e.g., a fine wire, the distinction between the external fringes and the internal fringes is no longer a striking one. When the bar is wide, the internal fringes lie extremely close together and may escape observation.

The foregoing statements regarding the spacing of the internal fringes may be theoretically deduced as follows:

It is shown in treatments on interference that the separation of the fringes in Young's experiment is:

$$z_0 = \frac{b\lambda}{w}, \tag{4}$$

where w is the distance between the two slits and b is the distance from the plane in which the slits are located to the plane of observation.

Passing to the consideration of our diffraction pattern, at the center there is an arc of length v' removed from each branch of the Cornu spiral. The radius of curvature of the spiral at the point located a distance v' from the origin, measured along the arc, is according to equation (2) of chapter 3 equal to:

$$R' = \frac{1}{\pi v'}. \tag{5}$$

Upon leaving the center of the pattern the points which correspond to O and O' of Figure 60a move, one clockwise and the other counter-clockwise, and after a quarter of a convolution the vectors which correspond to OJ and $J'O'$ are oppositely directed —the illumination is a minimum. After another quarter of a convolution, or after a semiconvolution from the starting-point, the vectors OJ and $J'O'$ again have the same direction and we attain the first maximum adjacent to the center of the pattern. Since the radius of curvature of the spiral is $R' = 1/\pi v'$, the length of a semiconvolution in the vicinity of the point located at a distance v' from the origin is:

$$v_0 = \pi R' = \frac{1}{v'}. \tag{6}$$

To the length v_0 measured along the spiral there corresponds a distance z_0 measured from the center of the pattern to the first adjacent maximum, the value of z_0 being obtained by applying equation (3), namely:

$$z_0 = v_0 \frac{a+b}{a} \sqrt{\frac{ab\lambda}{2(a+b)}}. \tag{7}$$

Substituting in accordance with equation (6), for v_0 the value $1/v'$ and then expressing v' by equation (2) in terms of q' and remembering that $2q' = w$, equation (4) is obtained for the separation of the internal fringes. This shows that the spacing does not depend upon the distance a, and is the same as in Young's interference experiment.

4. The Double Slit.—A typical pattern for a double slit is shown diagrammatically at the top of Figure 62. The pattern may be regarded as arising from the summation of two disturbances, one from each slit, the summation being carried out in accordance with the principle of interference, that is, taking due account of the difference of phase of each of the two disturbances. At the bottom of the figure the two patterns which would arise from each slit acting separately are represented. From considerations of symmetry it is clear that at the central field point, P_0, the two disturbances, one from each slit, must arrive in the same phase. Consequently

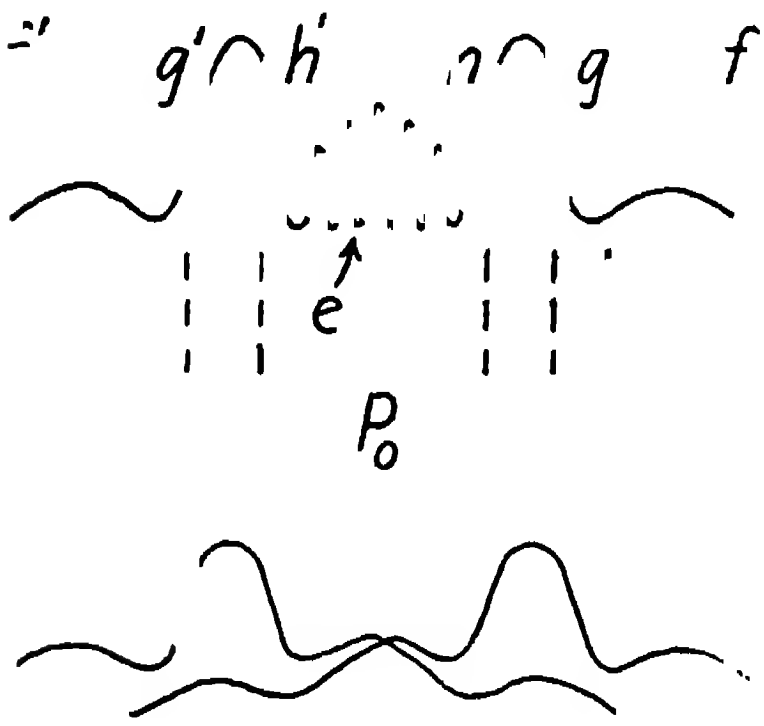


FIG. 62

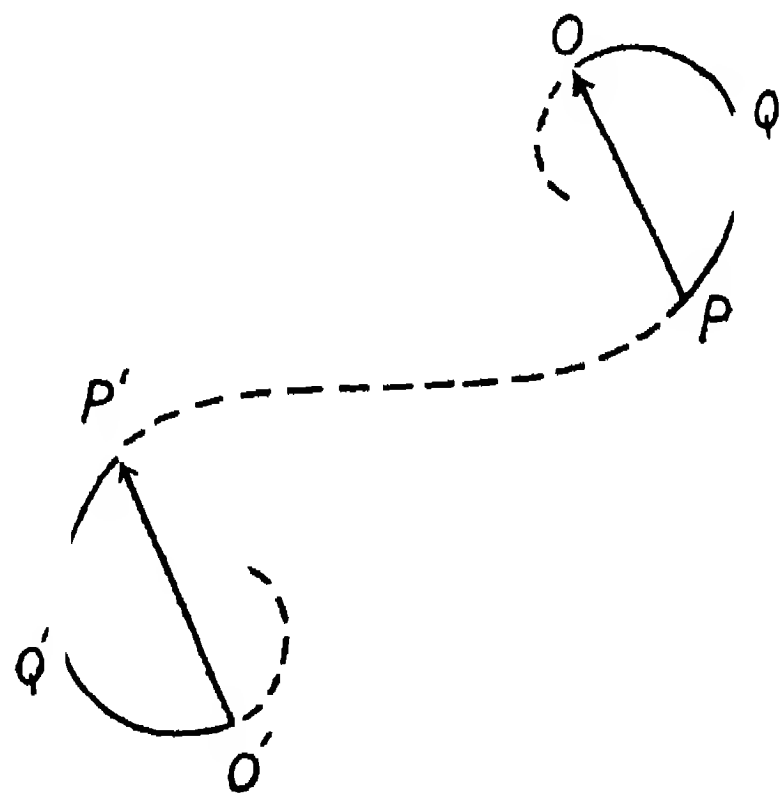


FIG. 63

at this point the height of the upper curve is equal to the sum of the heights of the lower curves, the ordinates of all three curves being taken to represent amplitudes. As we pass from the central point of the pattern to the left, the phase of the disturbance from the right-hand slit lags behind that from the left-hand slit and when the phase difference attains a value of π we arrive at a minimum of illumination (point e in the figure). Invoking the Cornu spiral, at the central point P_0 the arc $O'Q'P'$ (Fig. 63) is effective for the left-hand slit, and the arc OQP is effective for the right-hand slit. The closing vectors $O'P'$ and PO of these arcs are equal in length and are parallel —their resultant is obtained by taking their arithmetical sum—at the central point of the pattern. As we pass from the center to the left, the arc $O'Q'P'$ slides along the spiral in a clockwise sense and consequently the vector $O'P'$ rotates clockwise. The vector PO, on the other hand, rotates counter-clockwise. The situation resembles that which obtains within the shadow of a bar, and the reason for the occurrence of the closely spaced diffraction or interference fringes which are present in the region $h'h$ of the pattern (Fig. 62) should now be clear. In the regions hg and gf, and $h'g'$

and $g'f'$, similar, closely spaced fringes are not observed. The intensities of the two interfering beams differ greatly in these regions and the corresponding intensity variations are small in magnitude and escape observation.

There may be several broad maxima in which narrow fringes occur, between the two slits, instead of a single broad maximum as shown in our figure —depending upon how the conditions of the experiment are chosen. The reader who is interested should not fail to make his own observations. The slits should be two- or three-tenths of a millimeter wide and separated one or two millimeters, the distance from the source slit to the double slit being perhaps a meter. The slits may be conveniently made from short straight pieces of wire, separating these by inserting bits of metal of the required thickness between them, at their ends. By beginning the observations with the observing eyepiece or hand lens close to the slits and gradually withdrawing it, the pattern passes through beautiful and interesting transformations.

5. The Circular Aperture.—For the study of the diffraction pattern due to a circular aperture a point source of light is used. At a distance from the source which we designate by a, the diffracting screen is erected with the axis of the aperture passing through the source. The pattern is observed in a field plane beyond the screen at a distance b, the field plane being perpendicular to the axis of the aperture. The form of the resulting pattern depends markedly upon the diameter of the aperture; however, when the distances a and b and the wave-length λ are so chosen that each of two apertures subtends the same number of zones with reference to the axial field point, then the patterns are the same except for a scale factor. That this is true will become evident as we proceed.

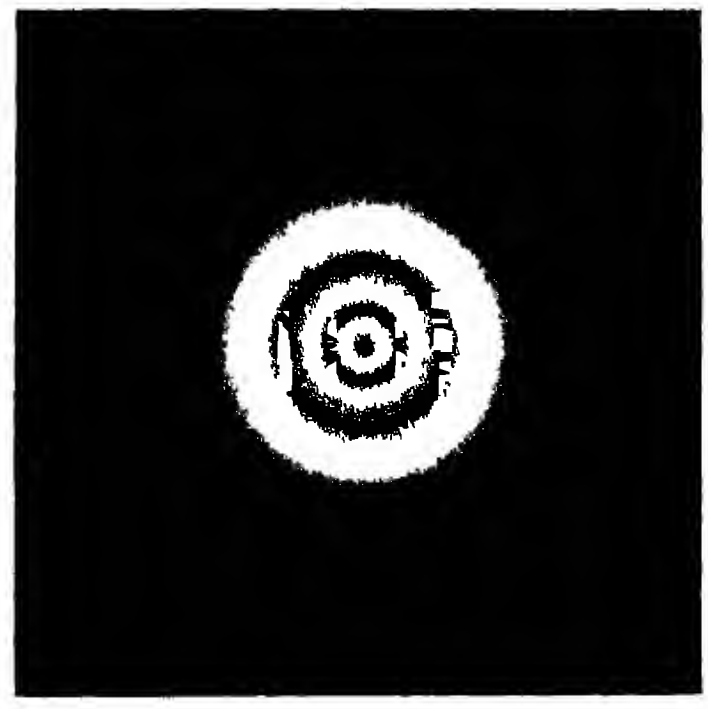


FIG. 64.—Circular aperture; six zones (from W. Arkadiew, *Phys. Zeit.*, **14**, 832, 1913).

In making observations the number of zones which a given aperture subtends may be varied by varying the distances a and b. The distance b may be varied continuously while looking through an observing eyepiece which is arranged to slide along an optical bench.

Figure 64 shows a photograph, by W. Arkadiew,* for an aperture which subtends six zones. Since the number of zones is an even one, the pattern has a dark center. In Figure 65, also by Arkadiew, the pattern for an aper-

* *Phys. Zeit.*, **14**, 832, 1913.

ture which subtends three zones is shown. The upper part of the figure shows an intensity curve calculated upon the basis of theory developed by E. Lommel.* The broken vertical and horizontal lines mark respectively the edge of the geometrical shadow and the intensity of normal illumination. As we should expect, the intensity at the center of the pattern is four times that of normal illumination. We may account qualitatively for the dark ring in the pattern and the surrounding brighter region: At the center of the pattern the effect of the second zone may be considered as annulled by the effect of either the first zone or the third, leaving the effect of one odd-numbered zone outstanding. Upon leaving the center of the pattern, and moving let us say to the left, the pole of the wave and the zone system move a corresponding distance to the left. The position of the aperture on the zone system is now as represented in Figure 66. A portion of the third zone is blocked off and a portion of the fourth zone is effective. The net effect of the portions of the various zones which pass is nearly zero, accounting for the low value of the illumination. At a point in the pattern farther to the left, a portion of the second zone, an even-numbered one, will be obstructed, and a portion of the fifth, an odd-numbered one, will pass. Consequently there is a brighter region beyond the dark region.

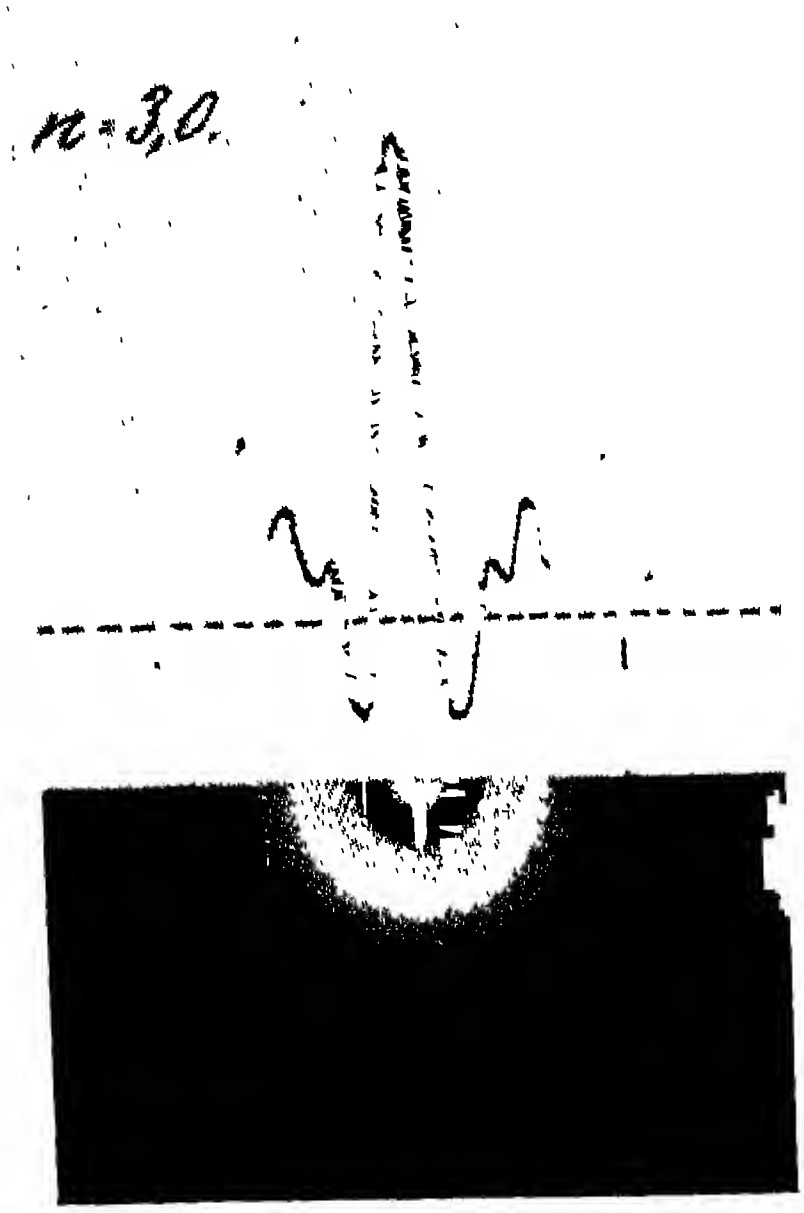


FIG. 65.—Circular aperture; three zones (Arkadiew).

The theory of the diffraction due to a circular aperture as given by Lommel involves extensive and difficult mathematics including a knowledge of Bessel's functions. This theory will not be given. Instead, a makeshift method of attack will be outlined in section 9.

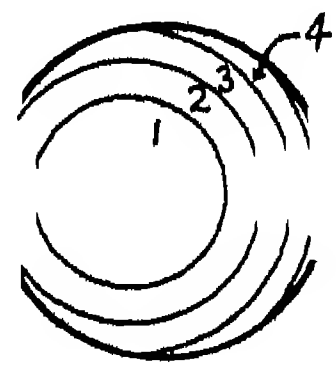


FIG. 66

* *Bayerisch. Akad. d. Wiss.*, **15**, 233, 1886, see also Gray and Mathews, *Treatise on Bessel's Functions*, chap. 14, or Bouasse et Carrière, *Diffraction* (Paris: Delagrave, 1923), chap. 13, esp. pp. 294 ff.

The pattern due to a circular aperture resembles the pattern due to a slit in so far as there is no well-marked separation of the fringes into two distinctly different sets, of which one falls outside of the geometrical shadow and the other falls within the shadow.

Recently Hufford and Davis* have applied the Lommel theory to apertures subtending relatively large numbers of zones. For this they had to extend the existing tables of Bessel's functions. They checked the results by photographing the corresponding patterns. Figure 67 is an example of

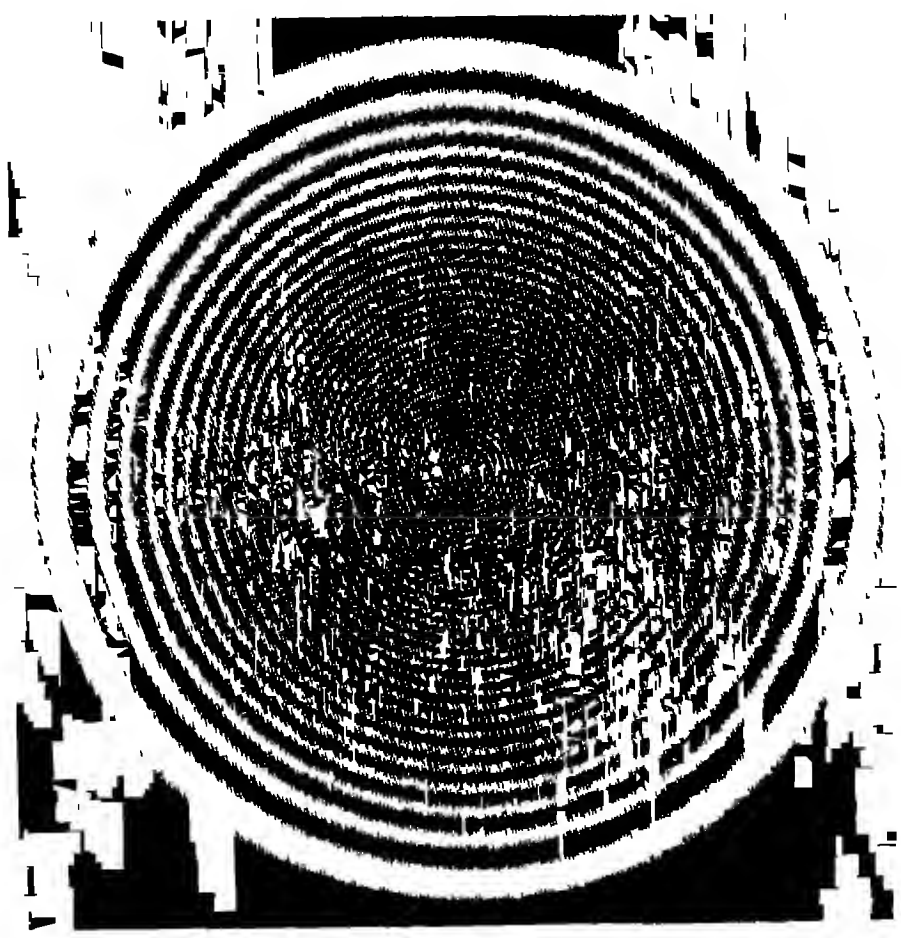


FIG. 67—Circular aperture subtending numerous zones (Hufford and Davis)

the photographs obtained by them. It is to be noted that the central spot is very small in diameter when the aperture is large.

The diameter of the central spot may be roughly calculated by considering the zone system to be moved from the central position across the aperture for a distance equal to the width of what is originally the outermost zone. This operation obstructs approximately half of the outermost zone and allows half of the next zone beyond to pass. Since zones of high number are narrower than those of low number, it follows that the diameter of the central spot is small when the diameter of the aperture is large and vice versa. And we may draw additional conclusions: We have thus far considered the source as a geometrical point; it is, however, in reality a small area. When the diffracting aperture is large, the diameter of the source must be especially small, otherwise the central region of the pattern will be obliterated by the overlapping of the patterns arising from different points in the source. Finally, the diffracting aperture must always be round with sufficient ac-

* *Phys. Rev.*, **33**, 589, 1929.

curacy—and the requisite perfection is greater for a large aperture than for a small one.

When the diffracting aperture is very small, subtending only a small fraction of a zone, then it approaches the ideal condition of constituting a point source from which Huygens' wavelets spread out in all directions. The central maximum of diffraction then extends far beyond the confines of the geometrical beam—only at field points lying far out in the field is there any possibility of destructive interference between disturbances arising from various points in the aperture. The central maximum spreads over a large area.

6. The Pinhole Camera.—The foregoing facts are of interest in connection with the pinhole camera. Referring to Figure 68, an object OO' may be photographed by means of a camera C which consists simply of a box having a pinhole H in one face. An image of the object is formed upon the photographic plate PP'.

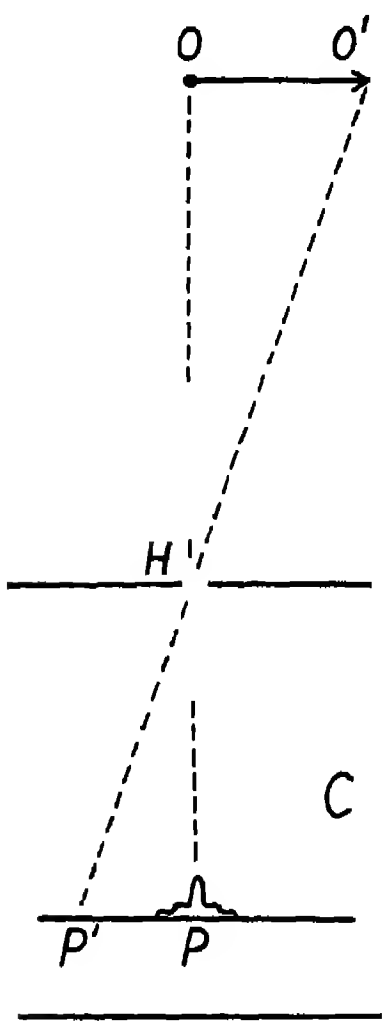


FIG. 68

The question arises: How large should the pinhole be in order to secure the best working of the camera? Each point of the object yields a diffraction pattern in the image. For example, the point O yields the pattern indicated at P. Obviously the hole H should not be very large, and if on the other hand it is too small, then the central maximum of diffraction will cover too wide a field. To obtain the best definition the central spot must be as intense and sharp as possible and the lateral maxima must be as faint as possible. When the intensity curves for apertures subtending various numbers of zones are examined, including those for fractional numbers, it becomes evident that the pinhole should subtend more than half a zone and less than one zone. The best value is about nine-tenths of a zone, because greater intensity is then obtained than when a value of let us say seven-tenths of a zone is chosen.*

7. The Circular Obstacle.—Examples of the patterns due to circular disks are shown in Figure 69, also from Arkadiew. When the disk obstructs several zones there are two distinct sets of fringes, an external set and an internal set, in analogy with the case of a straight bar. The center of the pattern is always a maximum of illumination. Moreover, the central spot becomes smaller in diameter as the disk becomes larger. The diameter of this spot may be calculated in the manner indicated in the case of the circular aperture. However, in the case of the disk when the source is so large

* Attention was first directed to this question by Lord Rayleigh, *Scientific Papers*, **3**, 429.

that it can no longer be properly considered as a point, then the central spot is not obliterated from the pattern by overlapping as in the case of the aperture. The central spot is the only very bright spot in the shadow—it will

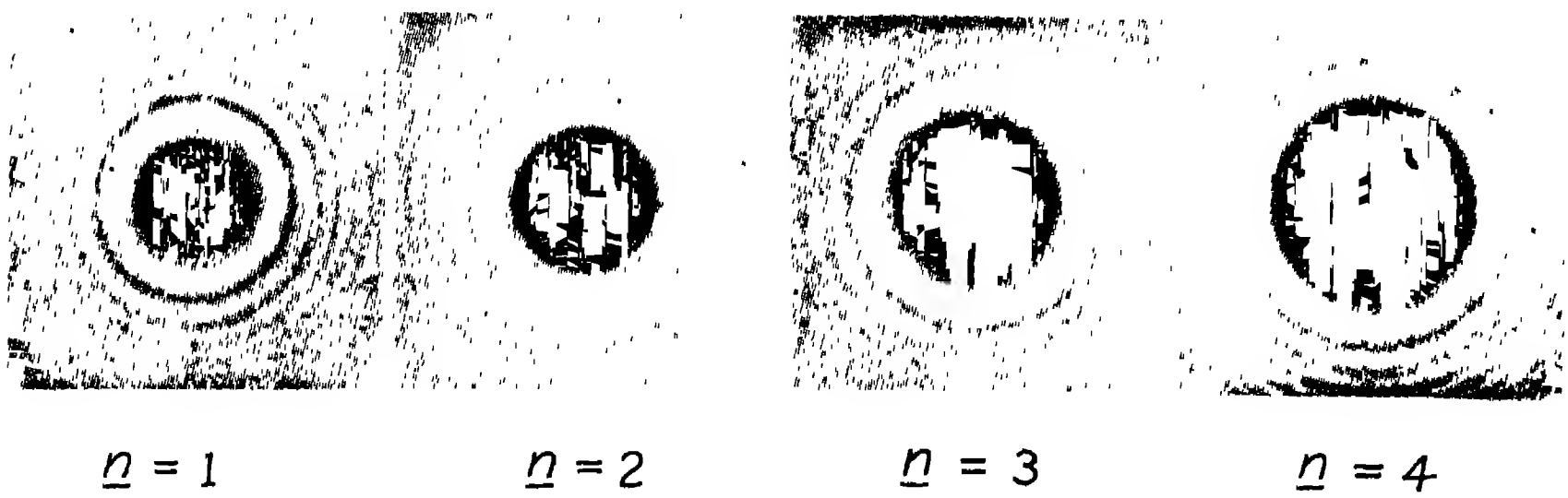


FIG. 69.—Circular disk (Arkadiew)

be recalled that a luminous object can be photographed around a disk. In the case of the disk there is a different though analogous effect when the source is too large to be properly treated as a geometrical point. This effect will be explained by reference to Figure 70. Suppose O and O' are points within the source separated perhaps by a distance of .2 mm. If the diffracting disk is large, the images P and P' of the points O and O' will be small and may be nearly or entirely separated from each other. However, at a point Q in the field plane, well outside of the shadow, the intensities due to the points O and O' of the source are additive. Consequently when the source has appreciable extension in area, the central spot behind the disk does not receive normal illumination, as it would if the source were a geometrical point—the central spot is relatively faint. It is principally for this reason that this patch cannot be readily found when the disk is large, especially when the pattern is viewed by allowing the shadow to fall on a sheet of white paper. In this case there is glare from the region surrounding the shadow. When the disk is large it is necessary to use an intense primary source such as an electric arc, and a *very small* pinhole as secondary source in order that there may be any hope of finding the central bright spot—and then it devolves upon us to find a *very minute* bright spot. Further, the disk must be round with sufficient accuracy. The various factors which have just been mentioned determine the success of the experiment—the decrease in effectiveness of the zones as we pass to those of higher number, owing to increased obliquity, is a factor which cannot

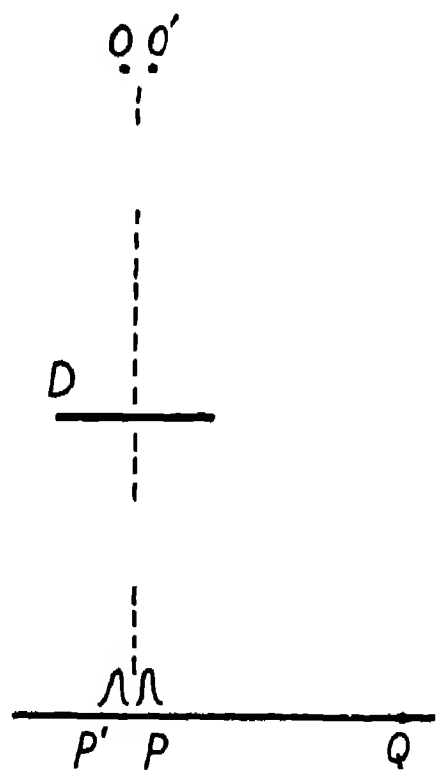


FIG. 70

reasonably be supposed to play an important rôle. As we pass to the use of larger and larger disks, or, what amounts to the same, observe in a field plane closer and closer to a given disk, the experiment fails long before the obliquity becomes appreciable—and it fails for the reasons above explained.

8. The Square Corner; the Tapered Slit.—A point source of light must be used.

Among the patterns observed by Grimaldi were those which arise from small obstacles with square corners. Figure 71 is a reproduction from his *De lumine*. This figure illustrates the fact that the *external* fringes bend

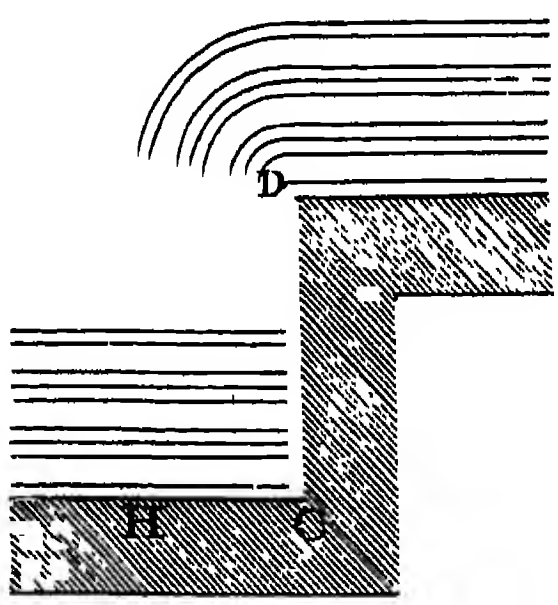


FIG. 71.—External fringes at square corners showing how the external fringes bend around an outside corner, D, and cross at an inside corner, C (from Grimaldi).

FIG. 72.—The tapered slit (from Bouasse et Carrière, *Diffraction* [Paris: Delagrave, 1923]).

around an outside corner, D, and cross at an inside corner, C. A figure, also from Grimaldi, showing the branching or crested *internal* fringes at a corner was reproduced in chapter 1. The fringes due to square corners are interesting to observe and they have a historic interest, but no theoretical interpretation of such fringes will be here attempted.

The pattern due to a tapered slit is also one which is interesting to observe and one which is of historic interest. Newton studied this pattern carefully and in his *Opticks* a drawing of the pattern will be found. Figure 72 shows a photograph of the pattern. The lines which form the inverted V mark the edge of the geometrical shadow. When the slit tapers gradually, the pattern, considered at any height, is practically the same as the pattern which would arise from a uniform slit which has the width of the tapered slit at this point. Thus the pattern due to a tapered slit presents in one picture the succession of patterns which would arise from a uniform slit of which the width is being varied. A noteworthy feature of the pattern is the

pronounced broadening of the central maximum, in the upper part of the figure, as the width of the slit approaches zero.

9. Graphical Methods for Circular Aperture and Obstacle.—The methods which are to be described will, without entering into involved mathematics, convey to the reader some idea of how the diffraction pattern due to a circular aperture or obstacle may be plotted. (The actual operations are tedious.)

Let us suppose that we have a circular aperture which subtends one zone: At the axial field point the vibration curve is a semicircle, curve A, Figure 73. This vi-

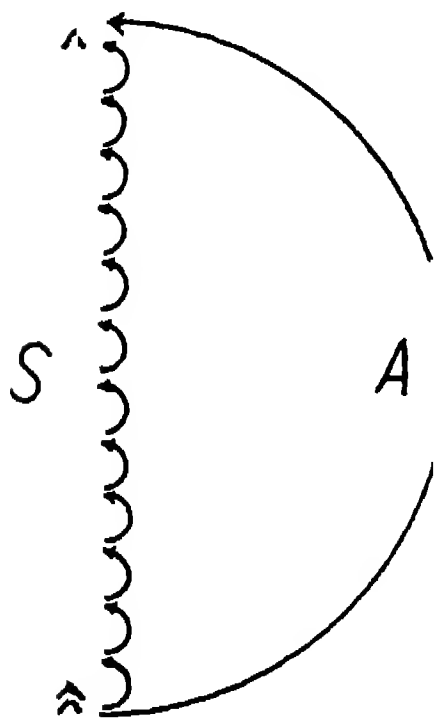


FIG. 73

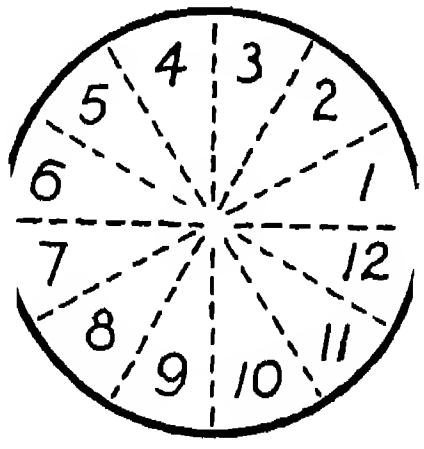


FIG. 74

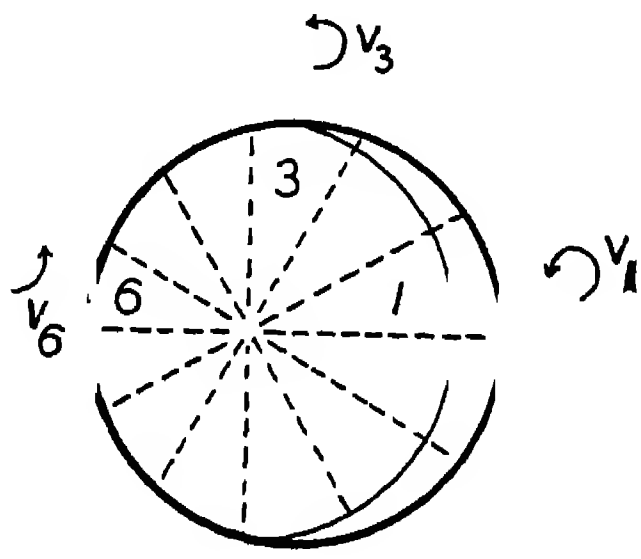


FIG. 75

bration curve has a closing vector S. We may divide the first zone into a number of sectors (see Fig. 74). The number of sectors will be designated by m; this number should be taken as twenty-four or more instead of twelve as shown. Each sector of the first zone will contribute a vibration curve which is a semicircle reduced m times in size in relation to the curve A. The closing vectors of these small semicircles yield a sum which is equal to the closing vector S of curve A. Moving from the center of the pattern, let us say to the left, the zone system will fall into the aperture as shown in Figure 75. Now, by way of example, sector No. 1 contributes the small vibration curve v_1, sector No. 3 contributes the curve v_3, and sector No. 6 contributes the curve v_6. From a graph of the zone system drawn accurately to a large scale, the magnitude and direction of the closing vector of the *vibration curve for each sector* are to be determined. Denote the closing vectors by s_1, s_2, etc. The sum of the vectors is now to be formed, namely:

$$S' = s_1 + s_2 + \cdots\cdots s_{12}$$

This is a vector sum. (Because of the existing symmetry it is really necessary to form the sum for only the sectors which fall, into let us say, the upper half of the zone system.) The square of the length of the vector S' is a measure of the intensity of illumination at the field point in question. This square should be obtained by computing the horizontal and the vertical components of each of the small vectors, then adding each set of components and thereupon squaring and adding these sums. The pattern may thus be plotted point by point.

(A curve actually plotted by this method for an aperture subtending three zones showed satisfactory agreement with the corresponding curve obtained by the Lommel theory given in Fig. 65.)

An alternative method is to divide each Fresnel zone into a sufficient number of minute zones (see Fig. 76). (This figure also applies to an aperture subtending one Fresnel zone.) The effect of each minute zone may be represented by a vector of which the direction is determined by the position of the minute zone in the zone system and the length represents the proportion of the minute zone which passes the aperture—this is to be ascertained by the use of a protractor. There is a vector for each minute zone, and this vector must be resolved into a horizontal and a vertical component. The corresponding components of the various vectors must be added—and the sums squared and added as before to obtain the intensity.

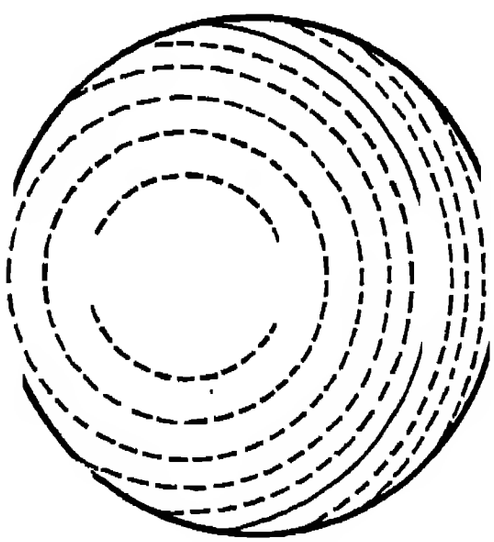


FIG. 76

For a circular obstacle, the portion of each minute zone which is *blocked off* by the obstacle must be determined and the effect of this portion represented by a vector. The various vectors so obtained must be added and the sum subtracted, vectorially, from the vector which represents normal illumination.

When an aperture and an obstacle subtend each the same number of zones, the major portion of the computation is the same in the two problems and hence does not require repetition. In actual application it is best to choose a case in which the aperture and obstacle subtend a *very* small number of zones. The number may be fractional; it should preferably not exceed two or two and one-half zones—the best case is that for one zone.

CHAPTER 5

THE FRAUNHOFER CLASS OF DIFFRACTION PHENOMENA

In our previous study of diffraction we have considered a source of light O (Fig. 77) which gives rise to a wave front MN, and we have then determined the effect which occurs at a given field point P. The path by way of a given point N of the wave front exceeds the path by way of the pole by the amount $\delta = NE$. The distance NE is determinable by describing an arc FE through the pole of the wave, about P as a center. Disturbances which occur at various points on the arc FE simultaneously will all reach P simultaneously. We shall refer to FE as a *diffracted wave front* with reference to the field point P. Or, more generally, we shall define a diffracted wave front with reference to a given field point as a surface of which all points lie at the same *optical* distance from the field point. A diffracted wave front is not really a wave front at all and must be distinguished from an *actual* wave front, for example, MN in our figure.*

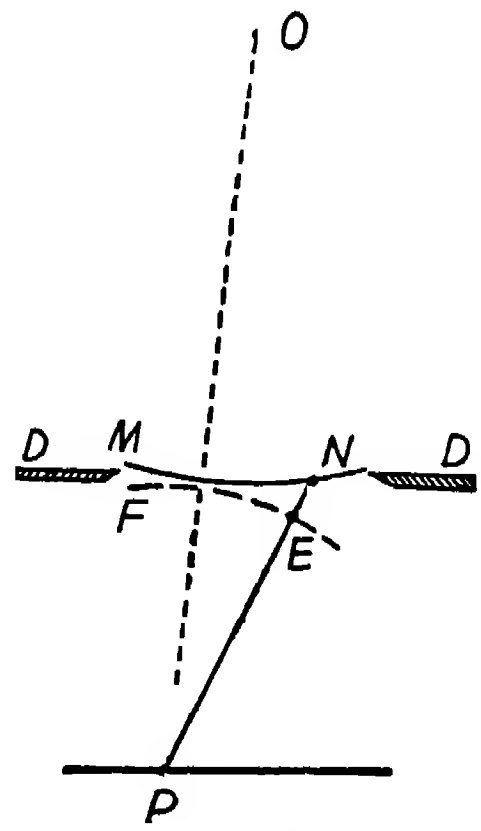


FIG. 77

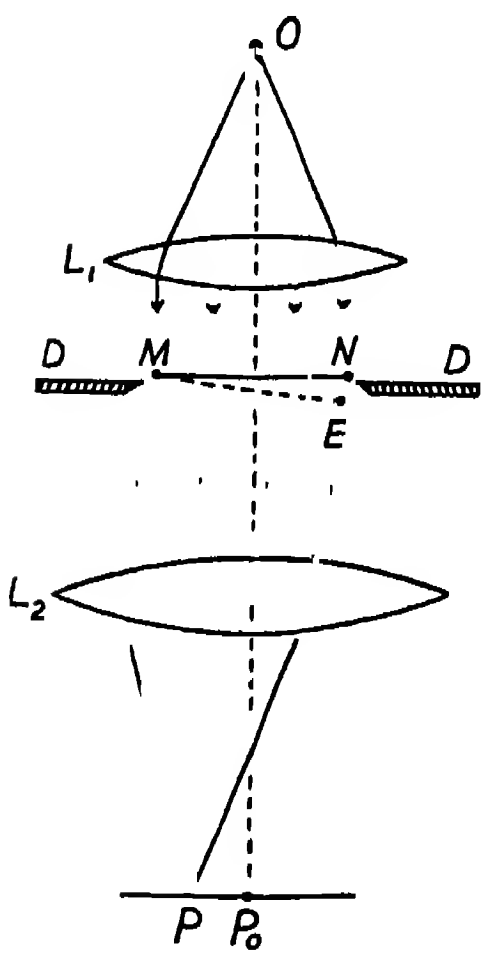


FIG. 78

It is possible by employing lenses to arrange matters so that both the actual and the diffracted wave fronts are *plane*, and it is under this condition that the important Fraunhofer class of diffraction phenomena can be most conveniently studied. These phenomena are of importance especially in connection with the theory of optical instruments.

1. Optical System.—Referring to Figure 78, suppose a point source of light, O, is placed at the principal focus of a lens, L_1. The actual wave fronts will then be plane after passing through this lens. Consider MN as a plane actual wave front, and let D be a diffracting screen. In all field planes beyond this screen, diffraction effects will occur. If we

* The term "diffracted wave front" is used with several different meanings. For the present we shall use it always with the meaning above stated.

choose to study these effects beyond a second lens, L_2, and in particular in the focal plane, PP_0, of this lens, then all points in a given plane ME will lie at a common optical distance from a certain field point P, and consequently all disturbances which occur simultaneously upon this plane will reach P simultaneously. Thus the plane ME is a diffracted wave front with reference to P and the idea of having a plane diffracted wave front is also realized.

It is to be assumed throughout that the lenses are perfect.

If it were not for diffraction there would be no illumination at all at the point P. There would be an image of the source O, formed in accordance with the laws of geometrical optics, at the point P_0, which is focally conjugate to O. The point P_0 is the only geometrically illuminated point in the field and is unique in this respect as well as in other respects of which we shall learn. It will be referred to as the central field point, or the central point of the diffraction pattern. These designations retain their meanings even when the diffracting aperture is not symmetrically placed with respect to the optical system or when indeed it has no symmetry. When the aperture in the diffracting screen D is large, or when there is no screen present, then the image at P_0 has a high degree of perfection, that is, it is small, and there is practically complete darkness at any point P fairly remote from P_0. When, on the other hand, the aperture in the screen D is small, then the image is broad and there may be considerable illumination at points fairly remote from P_0—well-marked diffraction effects occur.

2. Rectangular Aperture.—We shall suppose that the long edges of the aperture are vertical and shall discuss two cases: when a vertical slit is used as a source; and when a point source of light is used.

Case A. Vertical slit used as source.—This case is of especial interest in connection with the theory of the resolving power of the spectroscope.

The pattern consists of a system of vertical fringes which take the form represented diagrammatically in Figure 79. The central field point, P_0, re-

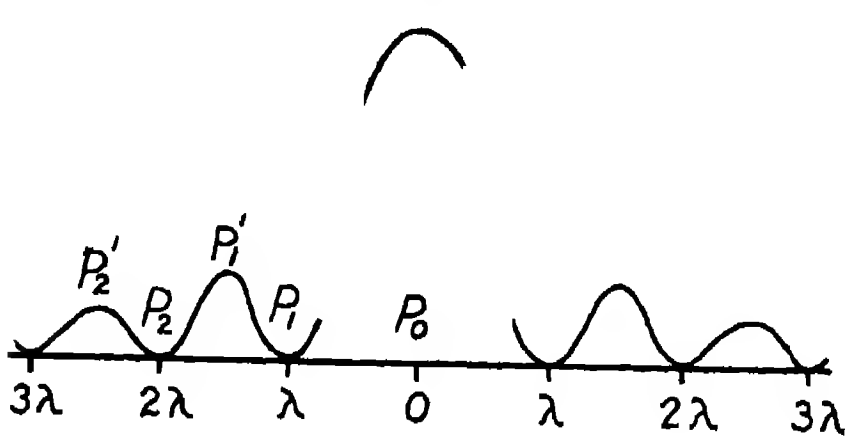


FIG. 79

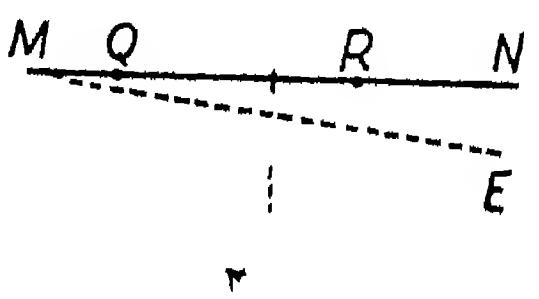


FIG. 80

ceives disturbances from all points of the aperture in the same phase, hence this point is always the most intensely illuminated point of the field. When we are at P_0, the diffracted wave front coincides with the actual wave front.

Let us now pass from P_0 to the left. Let MN (Fig. 80) represent the portion of the actual wave front which passes through the aperture. As we pass to the left the diffracted wave front is to be thought of as rotating about the point M as a hinge, becoming ME. Consider that the rotation has progressed until the path increment $\delta = NE$ takes the value λ. Let Q be a point in the left half of the aperture and let R be a point occupying a corresponding position in the right half of the aperture. The path by way of R now exceeds the path by way of Q by an amount $\lambda/2$. Hence disturbances traveling respectively by way of Q and R will annul each other. By pairing off points in similar fashion throughout the left half and the right half of the wave front we arrive at the conclusion that when $NE = \lambda$ the effects of the two halves of the wave will annul each other. Thus when $NE = \lambda$ we are at the first minimum of the pattern, at the point indicated by P_1 in Figure 79.

Proceeding farther to the left, when the path increment $\delta = NE$ takes the value $3\lambda/2$, we may divide the effective portion of the actual wave front into three equal parts. The effects of two of these parts which are adjacent will annul each other because the path difference by way of corresponding points in the two parts will again be $\lambda/2$. The effect of the third part is left outstanding. We are now at the first lateral maximum, indicated by P_1' (Fig. 79).

When NE takes the value 2λ, we may divide the wave front into quarters and conclude in similar fashion that the effects of adjacent quarters will annul each other, and hence the total effect will be zero. We are now at the second minimum, P_2.

When $\delta = NE$ takes the value $5\lambda/2$, we may divide the wave front into fifths. The effects of four of these fifths will annul each other and the effect of one-fifth will be left outstanding. We are now at the second lateral maximum, P_2'—and so forth.

The angle NME (Fig. 80) is equal to the angle between the axis of the optical system and the direction of diffraction. This angle is called the "angle of diffraction." The narrower the aperture, the greater will this angle be for a given value of the path increment NE, let us say for $NE = \lambda$, when the first minimum occurs. Thus, the narrower the aperture, the broader will the central maximum be—the broader will the *image* be.

A more thorough comprehension of the manner of formation of the pattern may be attained by considering the vibration diagrams which apply at various points:

Suppose Figure 81 represents the rectangular aperture. Divide this into a large number n of vertical strips of equal width as indicated. The number of strips should be thought of as being much larger than represented.

The vibration polygon for the central point of the pattern consists of n

vectors which all lie in a straight line (Fig. 82*a*), because the disturbances from the various strips all arrive at this point in the same phase. Moreover, the lengths of the vectors are all equal, because the strips all have equal areas. The closing vector of this straight polygon, drawn from the foot of the first vector to the head of the last one, is a measure of the amplitude of the illumination at the central point of the pattern.

FIG. 81

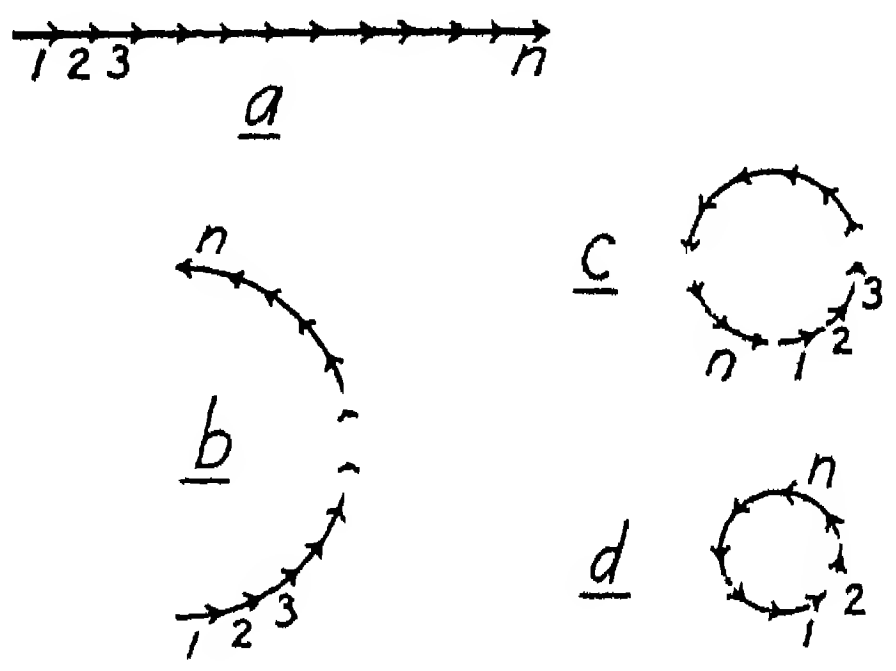


FIG. 82

As we leave the central point of the pattern and pass to the left, the vibration polygon curls up (see Fig. 82*b*), because the disturbance from each successive strip is now retarded with reference to the disturbance from the previous strip. The particular polygon shown in Figure 82*b* applies when the path increment δ, at the right-hand edge of the aperture, is $\lambda/2$, and Figure 82*c* applies when this path increment takes the value λ, or at the first minimum. The polygon here closes in one convolution. When the path increment takes the value $3\lambda/2$, the polygon makes one and one-half convolutions (Fig. 82*d*).

The polygons here shown are all both equilateral and equiangular. The closing vector drawn from the foot of the first vector to the head of the last one in each case measures the amplitude of the illumination which obtains at the field point in question.

By considering the number of vertical strips drawn on the wave front to increase indefinitely, while the widths of the strips decrease correspondingly, we may pass from the vibration polygons to corresponding vibration *curves*. These will all be sections of circular arcs.

In Figure 83 the diffraction pattern is represented to larger scale than before, and both an amplitude curve and an intensity curve are shown. Moreover, the vibration curves which apply at various key points of the pattern are inserted as detached curves.

(It will develop later that the positions of the maxima, determined as

above explained, are approximate positions. There is a slight correction to be made.)

We shall now determine the value of the ordinate of the amplitude curve at the various key positions shown in Figure 83. Let the vector A represent the vibration curve which obtains at the center of the pattern. The ampli-

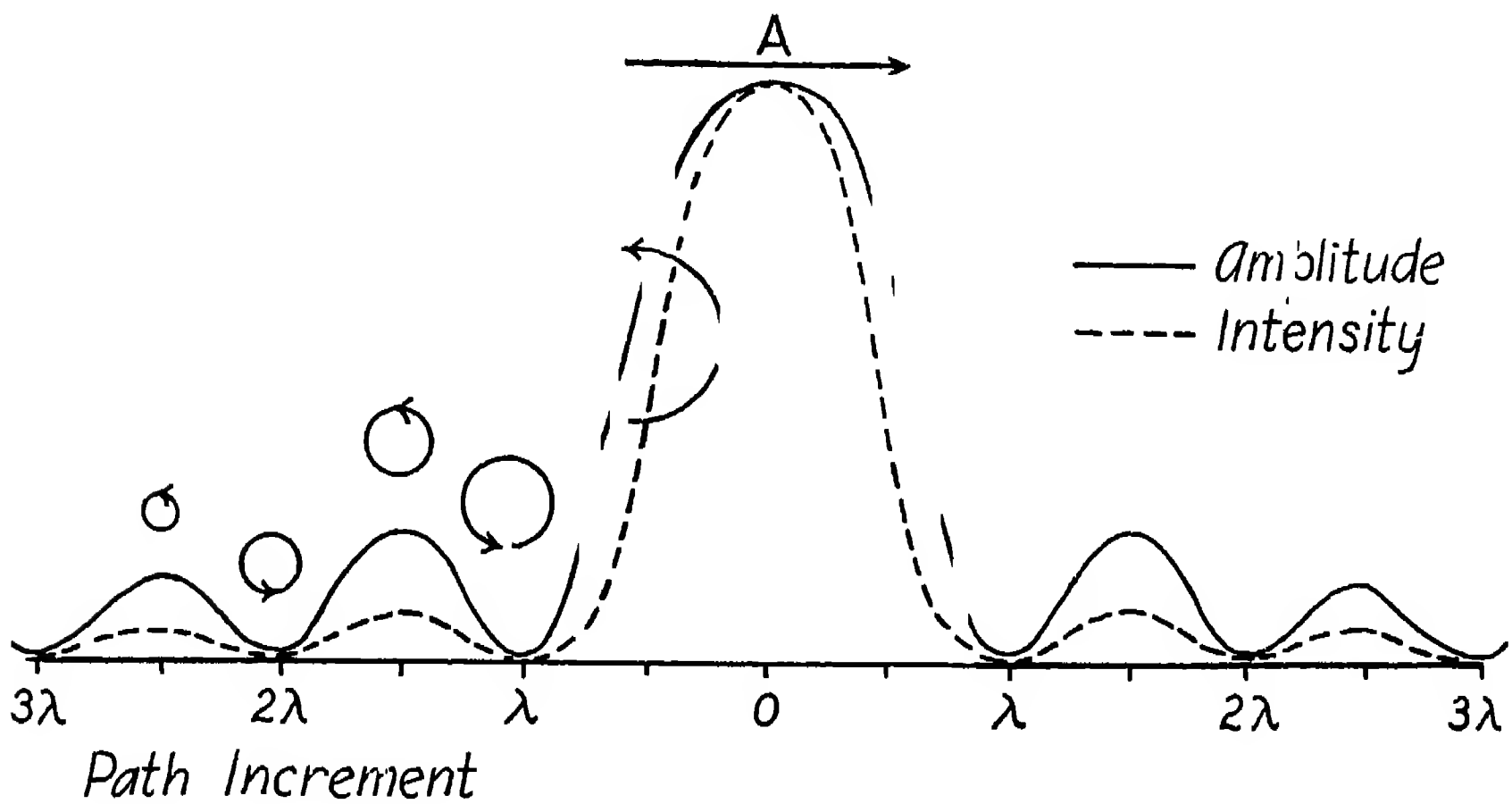


FIG. 83.—Rectangular aperture

tude of illumination at this point is thus to be represented by a magnitude A. The length A will also be the length of arc of the remaining vibration curves. Therefore, at the point halfway to the first minimum, where the vibration curve is a semicircle of length of arc A, the amplitude, the length of the closing vector, will be $2A/\pi$. The next key position is the first minimum. Here the ordinate is zero (and it is also zero at the succeeding minima). At the first lateral maximum the vibration curve is a circular arc of length A which has made one and one-half convolutions. The amplitude of illumination at this point is hence $2A/3\pi$. Similarly, at the second lateral maximum the amplitude is $2A/5\pi$—and so forth.*

In Fraunhofer patterns the central field point is always by far the most intensely illuminated point of the field. In contradistinction, in the Fresnel pattern, arising, for example, from a slit, the central point of the pattern may be either a maximum or a minimum depending upon the width of the slit. It is further to be noted that in the present case a single diagram suffices to portray the pattern for an aperture of any width. This is true because the distances from the center of the pattern to the various maxima and minima

* The ratio of the intensity of the first lateral maximum to the central maximum is $(2/3\pi)^2 = 1/22.2$, and the ratio of the intensity of the second lateral maximum to the central maximum is $(2/5\pi)^2 = 1/61.7$—and so forth.

bear a constant proportion to each other, regardless of the width of the aperture. Moreover, a single scale of abscissae suffices when this scale is laid off in terms of the path increment δ expressed in multiples of λ, as in our figure. If we wish to determine the linear distance, in the focal plane, from the center of the pattern to a given maximum or minimum, we proceed as follows: We apply the proportion $\delta/w = z/f$, where w is the width of the aperture, z is the desired distance measured in the focal plane, and f is the focal length of the focusing lens. In other words, $z = \delta f/w$. Now, for example, at the first minimum $\delta = \lambda$ and hence the distance of the first minimum from the center of the pattern is $z = \lambda f/w$.

Our treatment of the diffraction due to a rectangular aperture is as yet by no means complete, but it will be well, before continuing, to consider first some other matters. In particular it will be desirable to learn something of the formation of the Fraunhofer pattern due to a circular aperture and of the principles which govern the formation of Fraunhofer patterns in general. Accordingly, the present subject will be dropped for the time being in favor of these other matters. (The treatment of the diffraction due to a rectangular aperture is continued in sec. 6.)

3. The Fraunhofer Pattern Due to a Circular Aperture.— In optical instruments the aperture by which the light enters is commonly circular—the case which we are about to consider is of interest in the theory of the resolving power of the telescope and of the microscope.

The mathematical theory of the formation of the pattern was first given by G. B. Airy in 1834. His treatment is given in modified form in Appendix C, where, furthermore, references to the literature may be found. We shall content ourselves here with a qualitative explanation of the principal features of the pattern.

A point source of light is used. The pattern will obviously be symmetrical about the axial field point.

Consider the actual wave front to be divided into vertical strips of equal width. The portion of the actual wave front which falls within the aperture will then appear as represented in Figure 84. The number of strips should be thought of as much larger than represented. Let us construct a vibration polygon for each of various field points:

Taking first the central field point, the disturbances from the various strips all arrive there in the same phase, hence the polygon consists of vectors all lying in a straight line, Figure 85*a*. The lengths of the successive vectors are now by no means equal. The length of each vector is proportional to the area of the corresponding strip, or, since the strips have equal widths, the lengths of the vectors are proportional to the heights, or lengths, of the strips.

For a field point on the left of the central point so located that the path increment, δ, at the right-hand edge of the aperture is equal to $\lambda/2$, the polygon takes the form of Figure 85*b*. Each vector has the same length as before but is rotated by a constant amount with reference to its neighbors; the polygon, while far from being equilateral, is accurately equiangular be-

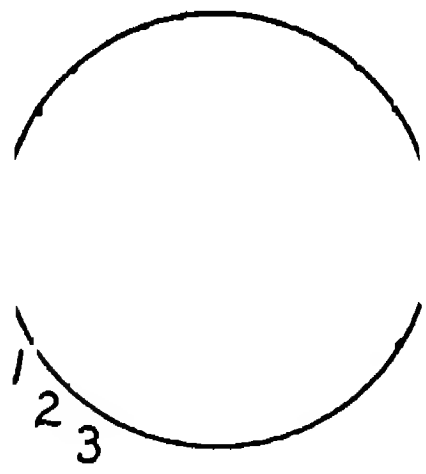


FIG. 84

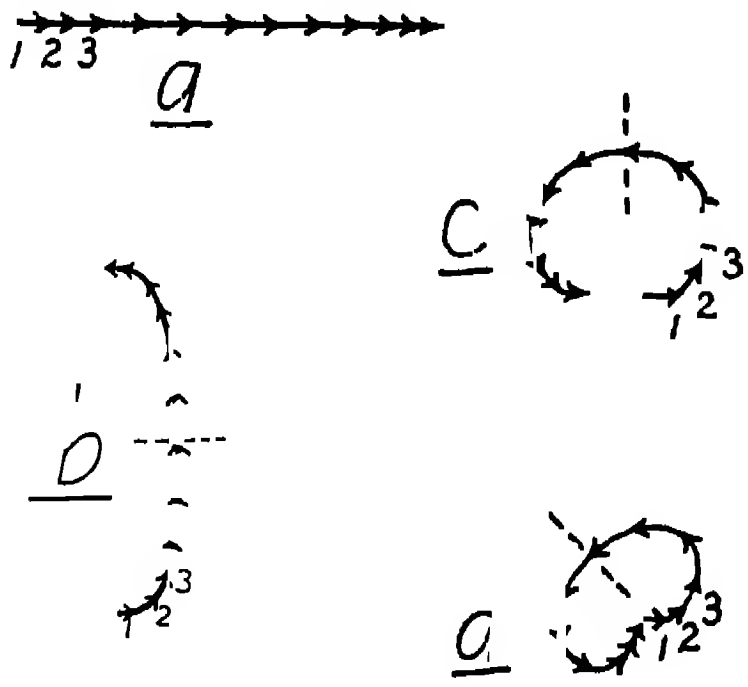


FIG. 85

cause the strips have equal widths and hence the phase retardation in passing from each strip to the next one is constant. Since in the case under consideration δ is equal to $\lambda/2$, the first vector and the last one point in opposite directions. The broken line is a line of symmetry.

The various vibration polygons for a circular aperture are each symmetrical about a perpendicular erected halfway along the polygon. The closing vector of each polygon measures the amplitude of the illumination.

For a field point so located that the extreme path difference δ is equal to λ, the vibration polygon takes the form of Figure 85*c*. It is again equiangular and now the first vector and the last one point in the same direction, horizontally to the right. The polygon makes one convolution but it does not close under this condition because it is not equilateral. We have not yet reached the first minimum in the pattern. The polygon must, however, close for a certain value of δ; this follows because the polygon is at all stages symmetrical. The mathematical treatment shows that the first closure (see Fig. 85*d*) occurs when $\delta = 1.22\lambda$. We are now at the first minimum. The illumination is zero here and at subsequent minima.

A diagram representing the pattern is given in Figure 86. A single diagram suffices for apertures of various diameters.

The first subsidiary maximum occurs somewhat *after* the vibration polygon—or, better, the vibration curve—has made one and one-half convolutions, and the second minimum occurs somewhat after the vibration curve has made two convolutions, etc.

The pattern consists of bright and dark rings surrounding the central maximum. Table II gives the positions in which these occur and also the corresponding values of the illumination.

The reader who desires to pursue the theory further is referred to Appendix C.

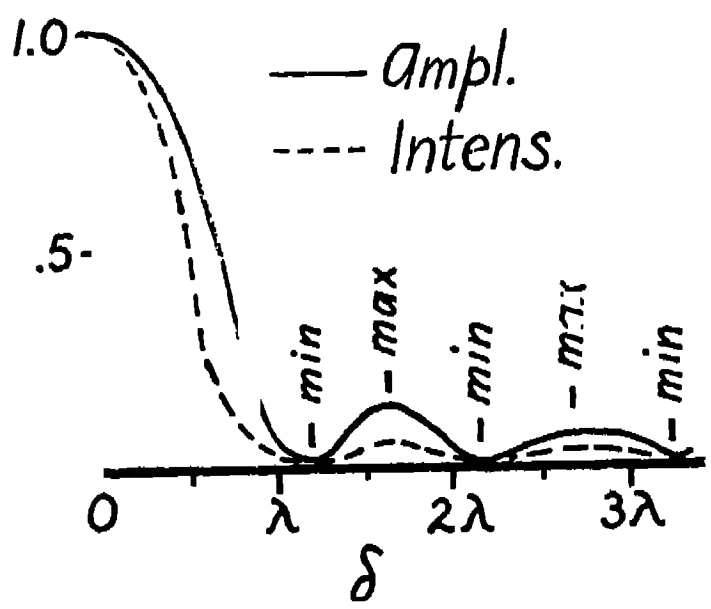


FIG. 86.—Circular aperture

TABLE II

CIRCULAR APERTURE

	Position, δ/λ	Amplitude	Intensity
Central max.	0	1	1
First min.	1.220	0	0
First submax.	1.635	0.1323	0.01750
Second min.	2.233	0	0
Second submax.	2.679	0.0615	0.00416
Third min.	3.238	0	0
Third submax.	3.699	0.0400	0.00160

4. General Principles of Fraunhofer Diffraction.—The general principles involved in the treatment of Fraunhofer diffraction phenomena will now be explained. These have already been applied in the foregoing discussion, but without much comment. We shall now call explicit attention to what they are:

Referring to Figure 87, let MN be the portion of the actual wave front which passes through any aperture and let ME be the diffracted wave front for a certain direction of diffraction indicated by the arrow. Let Q be a point on the actual wave front. The path by way of Q exceeds the path by way of M by the path increment $\delta = QT$. Consider Q as moving across the wave front from M to N. Denote the variable distance MQ measured across the wave front by q. We now have $\delta = QT = q \sin \theta$, where θ is the angle of diffraction. Thus in Fraunhofer diffraction the path increment δ is linearly proportional to q, the distance measured across the wave front. (In contradistinction, in Fresnel diffraction the path increment is proportional to the square of the distance measured across the wave front; see, e.g., eq. [5] of chap. 2.)

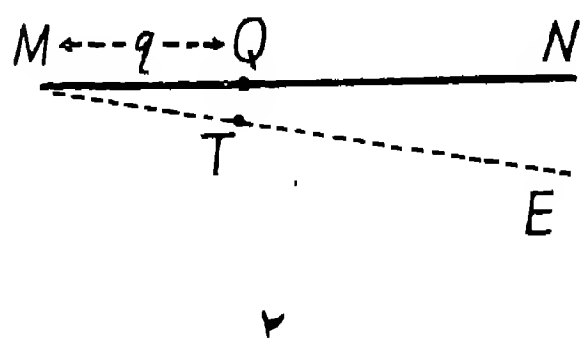


FIG. 87

The angle of diffraction θ is usually small, and when it is small we may equate the angle and its sine and its tangent. Denoting the focal length of the second lens by f, and distance measured from the center of the pattern

to a given point by z, we have $z/f = \tan\theta = \sin\theta = \theta$, and then, in view of the fact that $\delta = q \sin\theta$ as above explained, we have:

$$\delta = \frac{qz}{f}. \tag{1}$$

That is, for a field point lying at a distance of z cm from the center of the pattern, a disturbance traveling by way of a point lying q cm from the edge of the aperture will suffer a path increment δ as given by this equation. And the phase retardation with which this disturbance will arrive is $\psi = 2\pi\delta/\lambda = 2\pi qz/f$.

As the point Q crosses the wave front MN, the inclination of the direction of diffraction with reference to the element of the wave front where Q is located is throughout constant. Consequently there is no variation of the inclination factor to be considered when we draw the vibration polygon or curve which applies at a given field point. When the illumination at two different field points is compared, there is, strictly speaking, a difference in obliquity involved, but all obliquities which enter into question are small and consequently in the treatment of Fraunhofer diffraction phenomena consideration of the obliquity factor may be omitted and is, as a matter of fact, always omitted.

Referring now to Figure 88, let the contour of the figure represent an aperture of any shape. Consider an infinitesimally narrow vertical strip AB taken on the wave front. This strip is throughout its height or length at a constant optical distance from any given field point which lies on a horizontal line through the central field point. Thus the disturbances from all portions of the strip arrive simultaneously at any field point on this line. Denote the length or height of the strip by h and its width by dq, as indicated. Then the product $h\,dq$ is the area of the strip and is the measure of the amplitude contributed by the strip at the field point, or $h\,dq$ is to be taken as the measure of the length of the vector which represents the effect of this strip in any vibration polygon which may be drawn—and when the wave front is divided into strips of equal width, the lengths of the successive vectors are proportional to the heights, h, of the strips.

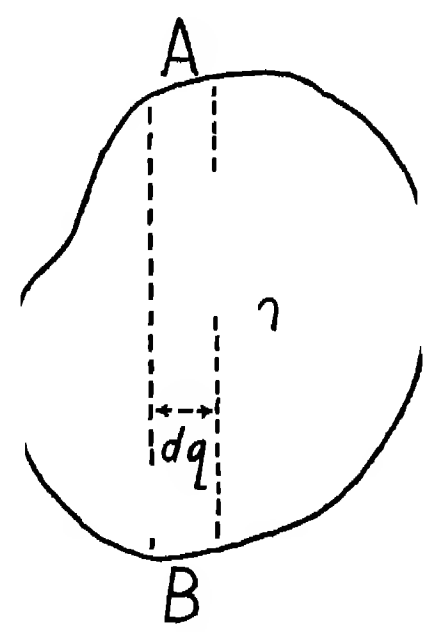


FIG. 88

In conclusion one more fact is to be noted: In the Fraunhofer arrangement the effective portion of the wave front is limited by the aperture of the optical system even when there is no diffracting screen present, and this circumstance entails two consequences. First, there is now no such thing

as "normal illumination." We now use the amplitude or intensity which obtains at the central field point as a basis of comparison in stating the illumination at other points. Second, there can be no Fraunhofer pattern which is analogous to the Fresnel pattern for a single straight edge or to the one for a bar. These diffracting systems require that the wave front be unlimited on one side or on both sides, or at least effectively unlimited. This condition cannot be fulfilled in the Fraunhofer arrangement. Therefore, in Fraunhofer diffraction the pattern due to any form of *aperture* or combination of apertures may come up for consideration, but not the pattern due to an *obstacle*.

5. The Fraunhofer Pattern for a Double Slit.—We shall now consider the pattern which arises from two narrow rectangular apertures—or, in other words, from two slits—which are supposedly of equal width. The slits are represented at the top of Figure 89 by MN and FH. We shall sup-

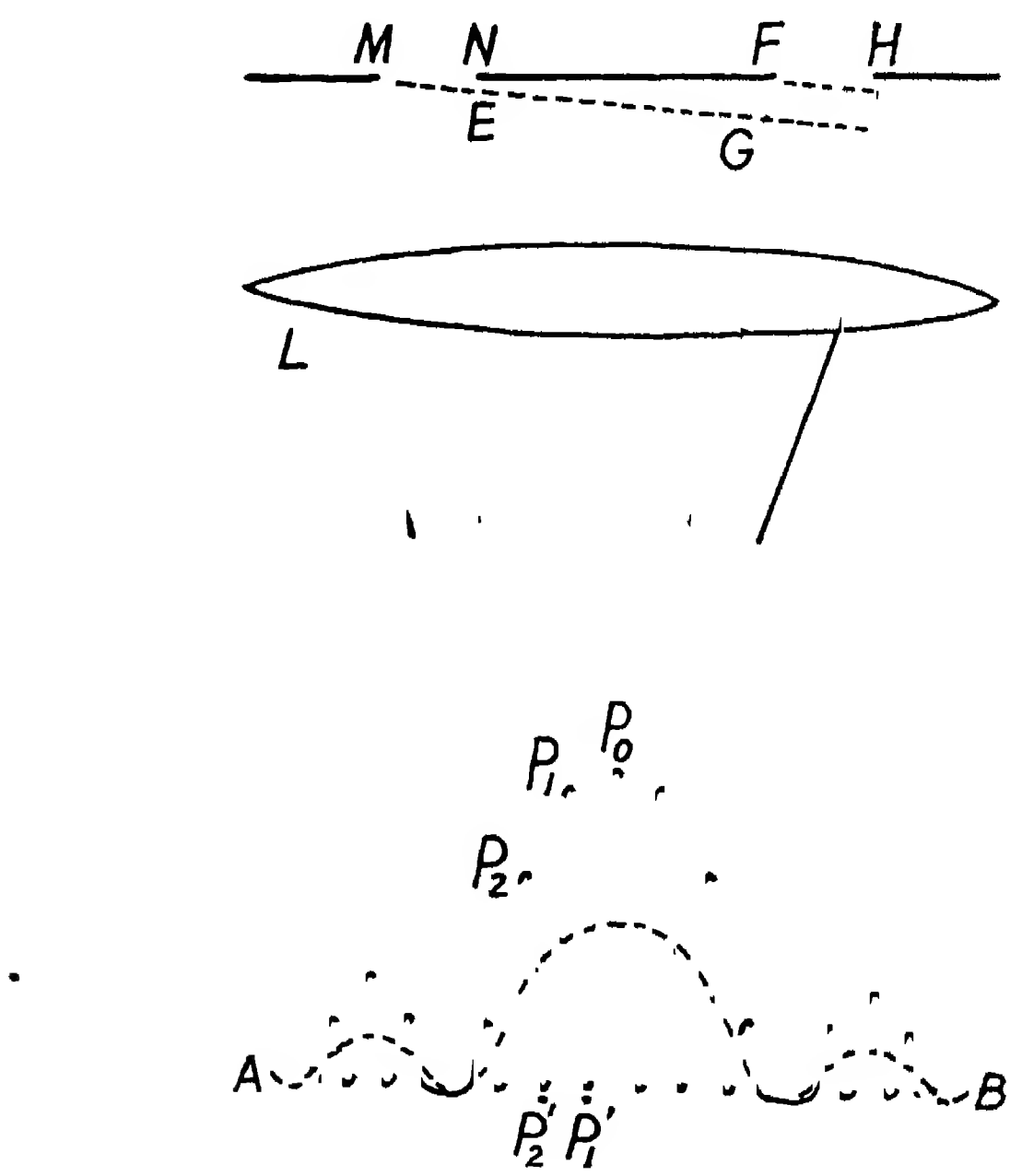


FIG. 89

pose that a linear vertical source of light is used and that the light, after being made parallel, strikes the diffracting screen from above in the figure. Below the screen, parallel diffracted rays, upon passing through the lens L, are brought to a focus in the plane of observation. The pattern will consist of fringes perpendicular to the plane of the diagram, or vertical. A diagram

of the pattern takes the form of the solid curve below, P_0, P_1, etc. Our problem is to determine why the pattern takes this form.

Let us suppose that the pattern from the left-hand slit acting alone is represented by the broken curve AB, amplitude being plotted vertically. Starting from the central point, we reach the first minimum on the left in the broken curve when the distance represented at the top of the figure by NE attains the value λ, as proved in section 2. This fact must be borne in mind. The right-hand slit acting alone would yield a pattern identical with that from the left-hand slit acting alone and, moreover, the two patterns would occupy exactly the same position in the field plane. The particular position of the diffracting aperture with reference to the aperture of the optical system affects neither the nature nor the position of the pattern. When the two slits act together, disturbances reach the central point of the pattern from both slits in the same phase, therefore the central maximum, P_0, is twice as high as the center of the broken curve. Upon leaving the center of the pattern and moving slightly to the left, the optical path from F exceeds that from M by the distance represented by FG. When FG attains the value of $\lambda/2$, the disturbances from M and F arrive out of phase, and so do the disturbances from any other pair of corresponding points of the two slits, for example, the disturbances from the two central points or those from the right-hand edges of the slits. Thus the phase difference from any pair of corresponding points equals the phase difference of the disturbances from the two slits each considered as a whole. Our problem thus reduces to one of determining the interference pattern for any pair of corresponding points and modulating this at every point in proportion with the illumination which each separate slit sends to the point in question. To continue, when the distance FG attains the value $\lambda/2$, as above supposed, the value of NE will be only a small fraction of a wave-length, therefore the first minimum in the solid curve, P_1', will occur long before the first minimum in the broken curve is reached. The first lateral maximum in the solid curve, P_1, will occur when $FG=\lambda$, for then the disturbances from F and M will reinforce each other as will also the disturbances from other pairs of corresponding points of the two slits. The second minimum, P_2', will occur when $FG=3\lambda/2$; the second lateral maximum, P_2, will occur when $FG=2\lambda$, etc. When FG attains a value so great that $NE=\lambda$, then the disturbance from each slit separately is zero, therefore the solid curve must have a minimum where the broken curve has one. Beyond this common minimum, interference again gives rise to closely and uniformly spaced maxima. Throughout, each maximum is twice as high as the broken curve at the same point.

The maxima of the broken curve are called maxima of the *first class*—

they arise from a single aperture. Maxima which arise from two or more apertures, for example, P_1, P_2, etc., are called maxima of the *second class*.

The vibration diagrams for several points of the pattern are shown in Figure 90. This figure should be carefully examined. The vibration curves for the separate slits are drawn in solid line. Each of the two arcs maintains its length at various points of the pattern and the arc represented in broken line between the two vibration curves likewise maintains its length throughout. The lengths of the closing vectors vary, however, from position to position. The total effect at any point of the pattern is obtained by vectorial addition of the closing vectors of the two vibration curves, one for each slit. At a certain stage beyond any of the stages represented, each separate vibration curve will close in one convolution. The effect of each separate slit will then be zero. This is the stage which obtains at the first minimum of the broken curve AB of the previous figure.

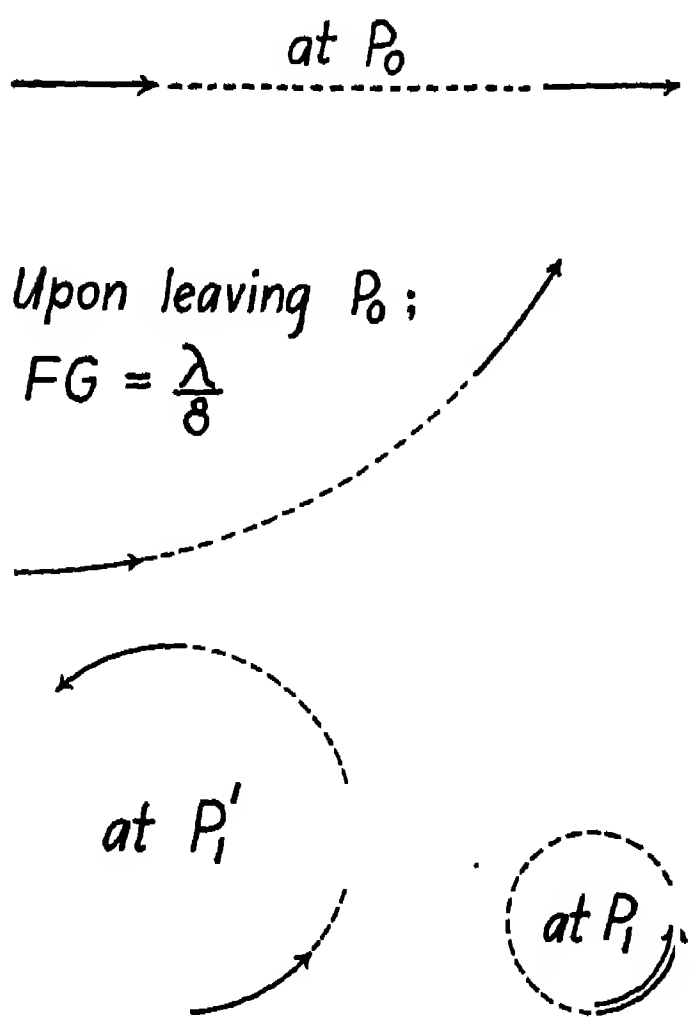


FIG. 90

When the pattern is observed visually through an eyepiece, it is a simple matter to pass continuously from the Fraunhofer pattern for a double slit to the corresponding Fresnel pattern. It is only necessary to slide the viewing eyepiece along an optical bench toward the diffracting system, the observer's eye being held always at the eyepiece. The ensuing changes should be carefully noted. The Fresnel pattern passes through various stages. Figure 91 shows, at the top, the Fraunhofer pattern, and below, two stages of the Fresnel pattern. The first of these stages was previously considered (see chap. 4, Fig. 62).

It has been aimed in the foregoing sections to present the salient features of the theory of Fraunhofer diffraction phenomena. The development of the theory is continued in the remainder of the present chapter. But the reader who is pressed for time may well pass on immediately to the study of the grating, in the next chapter.

6. Rectangular Aperture (***Concluded***).—The Fraunhofer diffraction pattern for a rectangular aperture, when a slit is used as the source, takes the form represented diagrammatically in our previous Figure 83. We shall now derive an equation from which any desired number of points of the curves of this figure may be plotted. The treatment will refer in particular

to the amplitude curve, the intensity curve being derivable from this by squaring the amplitude at each point.

Consider the ordinate of the amplitude curve at any given point. The value of this ordinate is determined by the length of a closing vector S (Fig. 92) of a circular arc of length A. The angle ψ subtended by this arc represents the difference in phase between disturbances reaching the field point under consideration by way of the two opposite edges of the aperture. Denote the radius of the arc by the symbol ρ, then $A/\rho=\psi$ and $(S/2)/\rho=\sin\psi/2$. By eliminating ρ we have:

$$S=\frac{A\sin\frac{\psi}{2}}{\frac{\psi}{2}}. \tag{2}$$

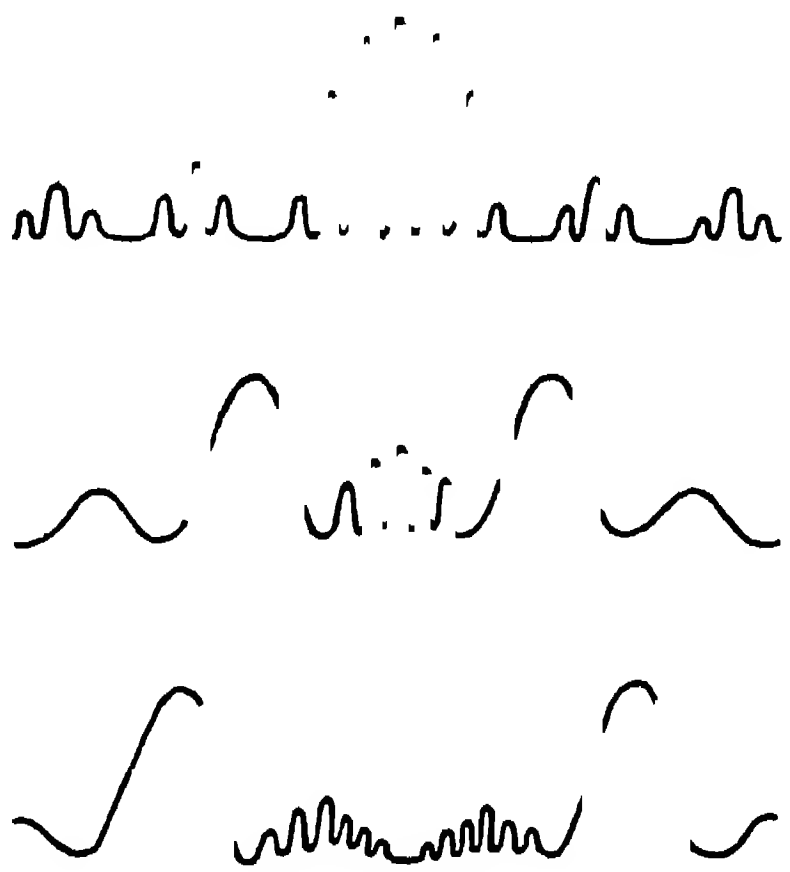


FIG. 91

This is the desired expression; from it any number of points of the curve may be plotted, bearing in mind that $\psi=2\pi\delta/\lambda$, where δ is the path increment at the edge of the aperture.

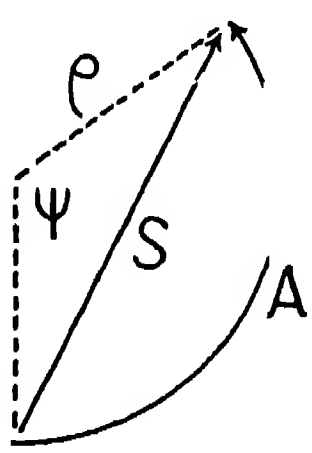


FIG. 92

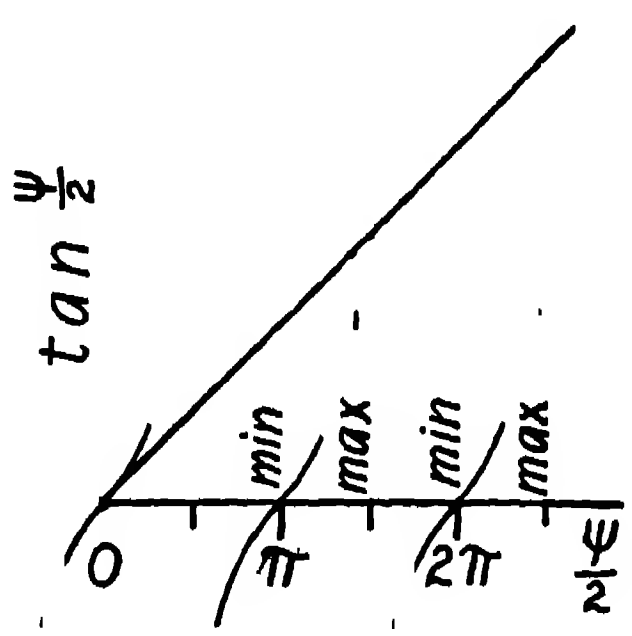


FIG. 93

The maxima of the curve may be located by differentiating equation (2), thus:

$$\frac{dS}{d\psi}=\frac{A}{\psi}\left[\cos\frac{\psi}{2}-\frac{\sin\frac{\psi}{2}}{\frac{\psi}{2}}\right]=0, \tag{3}$$

whence:

$$\tan \frac{\psi}{2} = \frac{\psi}{2}. \tag{4}$$

Figure 93 represents a series of tangent curves. Values of $\psi/2$ are plotted as abscissae and $\tan \psi/2$ as ordinates. A straight line through the origin which has a slope of 45° has its ordinates equal to its abscissae, consequently its intersections with these tangent curves satisfy equation (4) and thus give the positions of the maxima. The locations of the intersections are given in Table III, together with the approximate positions of the maxima

TABLE III

POSITIONS OF MAXIMA

Approximate		Exact
δ	$\psi/2$	$\psi/2$
$3\lambda/2$	1.50π	1.43π
$5\lambda/2$	2.50π	2.46π
$7\lambda/2$	3.50π	3.47π
$9\lambda/2$	4.50π	4.48π

which we previously determined by considering that the maxima occurred when the vibration curve has made respectively one and a half convolutions, two and a half convolutions, etc. The first maximum occurs somewhat before the vibration curve has made one and a half convolutions, because this curve while it winds up also contracts in diameter. The second maximum occurs, to a less degree, before the vibration curve has made two and a half convolutions. As we pass progressively to maxima of higher number, the approximate positions progressively approach the exact ones.

The positions of the minima which we determined as occurring when the vibration curve makes exact integral numbers of convolutions are precise and require no correction.

Case B. Point source of light.—The pattern which arises from a rectangular aperture when a point source of light is used is shown in Figure 94. The aperture itself is shown as an inset in the lower right-hand corner of the figure. Let us consider how this pattern arises.

Along the central horizontal line of the pattern the illumination varies in exactly the same way as it varies in crossing the vertical fringes which are formed when a slit source of light is used, as in Case A. The theory which was given for that case applies verbatim for the illumination along the central horizontal line. Along the central vertical line the illumination also varies according to the same law, but the scale is now different. In consider-

ing the illumination along this line the aperture should be thought of as being divided into horizontal strips, as illustrated in Figure 95.

The central rectangular maximum of the pattern has the same proportion of length to breadth as the aperture, only it is laid over on its side. This will be proved: Referring to Figure 96, suppose P_0Z represents the central horizontal line and P_0Y the central vertical line. The first minimum along the

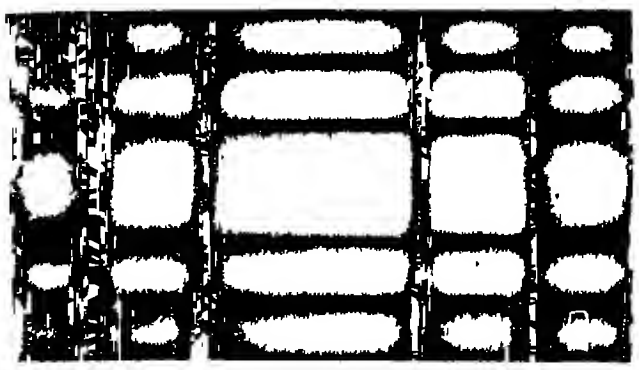


FIG. 94.—Pattern due to rectangular aperture; point source of light (Lummer und Reiche).

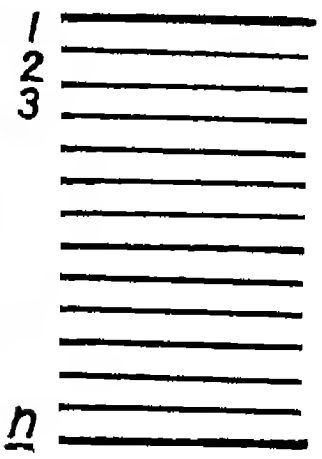


FIG. 95

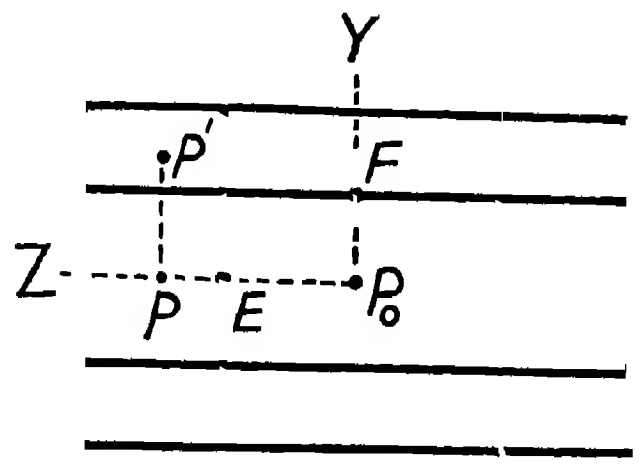


FIG. 96

horizontal line occurs, let us say, at E, which lies at the co-ordinate $z = f\lambda/w$, where f is the focal length of the second lens and w is the width of the aperture. The first minimum along the central vertical line occurs at F, which lies at the co-ordinate $y = f\lambda/h$, where h is the height of the aperture. The central maximum, therefore, has the proportion of width to height $z/y = h/w$, which is the proportion of the height of the aperture to its width.

We shall now derive a general equation for the amplitude of illumination, and then account for the remaining features of the pattern:

Consider the aperture to be divided into a large number, n, of horizontal strips, as in Figure 95. If the amplitude of illumination at the center of the pattern due to the whole aperture is denoted by A, then the amplitude contributed by one horizontal strip at the center of the pattern is $A_1 = A/n$.

Let us now proceed from the central point of the pattern horizontally to the left. As we proceed, the diffracted wave front turns about the left edge of the aperture as a hinge. Stopping at some point, P (Fig. 96), consider the vibration curve for one horizontal strip. The length of arc of this vibration curve is A/n, and we may determine the length of the closing vector of this curve, which we shall denote by S_1, by substituting the value A/n in equation (2) for A. Thus:

$$S_1 = \frac{\frac{A}{n} \sin \frac{\psi}{2}}{\frac{\psi}{2}}. \tag{2'}$$

There is one vector of length S_1 for each horizontal strip. While we are at the field point P, these vectors, to the number of n, all have the same direction—they may be considered to form a straight line of length nS_1. The diffracted wave front for the point P coincides along its left edge with the edge of the aperture and extends forward on the right, like a door slightly ajar toward us. Directing our attention to the upper horizontal edge of this diffracted wave front, when we proceed from the field point P vertically upward, this edge remains fixed. It acts as a hinge for the diffracted wave front which now swings forward at the bottom. The disturbance, of amplitude S_1, from each successive lower horizontal strip is now retarded in phase with reference to the disturbance from the strip above. Our vectors, each of length S_1, and n in number, form a vibration polygon, or, passing to the limit, a vibration curve having a length of arc nS_1, and we can determine the length of the closing vector of this vibration curve for any point P' (Fig. 96) lying vertically above P. The length of this closing vector will represent the effect of the entire aperture—it will represent the amplitude of the illumination which obtains at *any* given point P' of the pattern (because P may be chosen to lie anywhere along the horizontal axis). Applying equation (2) again, the length of arc of our curve is nS_1, and this value is to be substituted for A in this equation. For ψ in this equation we shall write ψ' to indicate that we are now dealing with the difference of phase between disturbances traveling one by way of any given point on the upper edge of the aperture and the other by way of a point on the lower edge of the aperture vertically below the first point. We thus have for the amplitude of illumination at any point P' of the pattern:

$$S=\frac{nS_1 \sin \frac{\psi'}{2}}{\frac{\psi'}{2}},$$

or, inserting the value of S_1 from equation (2′):

$$S=A\,\frac{\sin \frac{\psi}{2}}{\frac{\psi}{2}}\,\frac{\sin \frac{\psi'}{2}}{\frac{\psi'}{2}}. \tag{2''}$$

This equation gives the amplitude at any point of the pattern. The intensity is proportional to S^2.

The intensity at a given point P' of the pattern is equal to the product of the intensities of the points which lie opposite it on the central horizontal and vertical lines. Thus the rectangular patches which lie along diagonal lines are decidedly fainter than those which lie along the central horizontal

and vertical lines. The diagonal patches are in fact so faint that they can be well observed visually only when the illumination from the brighter portions of the pattern is screened off.

The vertical lines of zero illumination occur where $\sin \psi/2=0$—or in other words, where ψ takes one of the values 2π, 4π, 6π, etc. We have in general $\psi=2\pi\delta/\lambda$, where δ is the path difference by way of two points, one lying anywhere along the left edge of the aperture and the other lying along the right edge at the same height as the first point. Moreover, we have $z=f\delta/w$, where z represents the abscissa of the vertical line in the field plane for which the corresponding extreme path difference is δ. Accordingly, the vertical lines of zero illumination occur where:

$$\psi=2\pi\,,\ 4\pi\ ,\quad \ldots .$$

$$\delta=\lambda\,,\ 2\lambda\,,\quad \ldots .$$

$$z=\frac{f\lambda}{w}\,,\ \frac{2f\lambda}{w}\,,\quad \ldots .$$

Similarly, the horizontal lines of zero illumination occur where ψ' and δ' take respectively the same series of values as ψ and δ and the ordinates take the values:

$$y=\frac{f\lambda}{h}\,,\qquad \frac{2f\lambda}{h}\,,\quad \ldots .$$

There is illumination at all other points of the pattern.

The centers of the rectangles which are formed by the vertical and horizontal lines of zero illumination have co-ordinates z and y obtained by combining two of the following:

$$z=\frac{3f\lambda}{2w}\,,\qquad \frac{5f\lambda}{2w}\,,\quad \ldots .$$

$$y=\frac{3f\lambda}{2h}\,,\qquad \frac{5f\lambda}{2h}\,,\quad \ldots .$$

The maxima of illumination occur at points slightly displaced from the centers of the rectangles—toward the central point of the pattern. This follows from the fact that the maxima of illumination along the central horizontal and vertical lines do not occur strictly halfway between the minima but lie somewhat closer to the center of the pattern, as exhibited in Table III.

7. Distinctions between Fraunhofer and Fresnel Diffraction.—If, in the optical system for securing Fraunhofer diffraction, the source of light were removed to infinity, we could dispense with the first lens and still have

plane waves incident upon the diffracting aperture. In addition, when the focal length of the second lens is great, then the plane in which the pattern is observed lies far removed from the lens. If the focal length of this lens were infinite, which is the equivalent of having no lens at all, then the field plane would lie at infinity. Accordingly, it is often said that Fraunhofer diffraction may be considered as that special case of diffraction without lenses, Fresnel diffraction, in which both the source of light and the plane of observation lie at infinity. However this statement, while correct, leaves the impression that Fraunhofer diffraction phenomena can be obtained only when the actual and the diffracted wave fronts are both plane, but such is not the case.

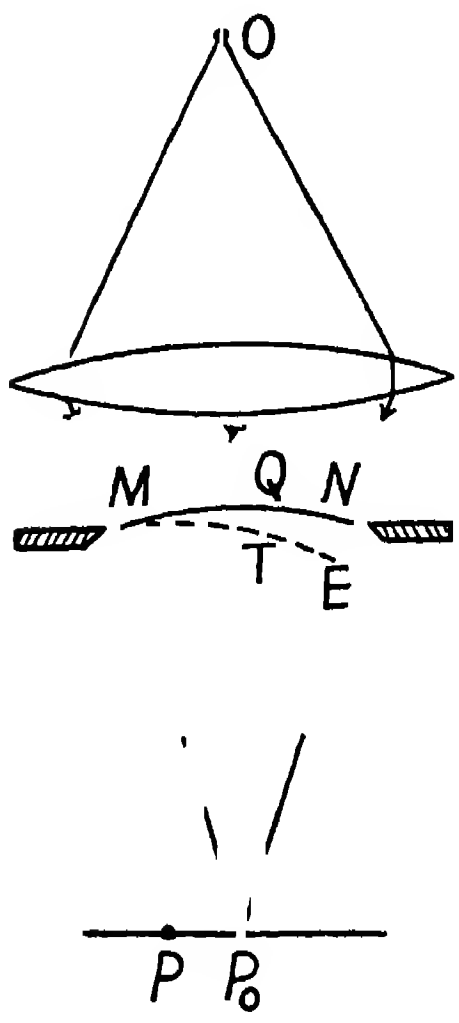


FIG. 97

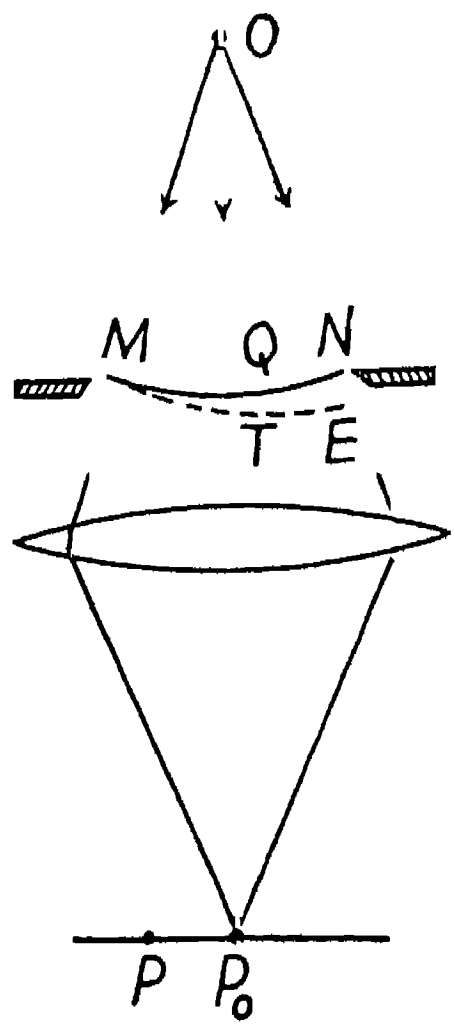


FIG. 98

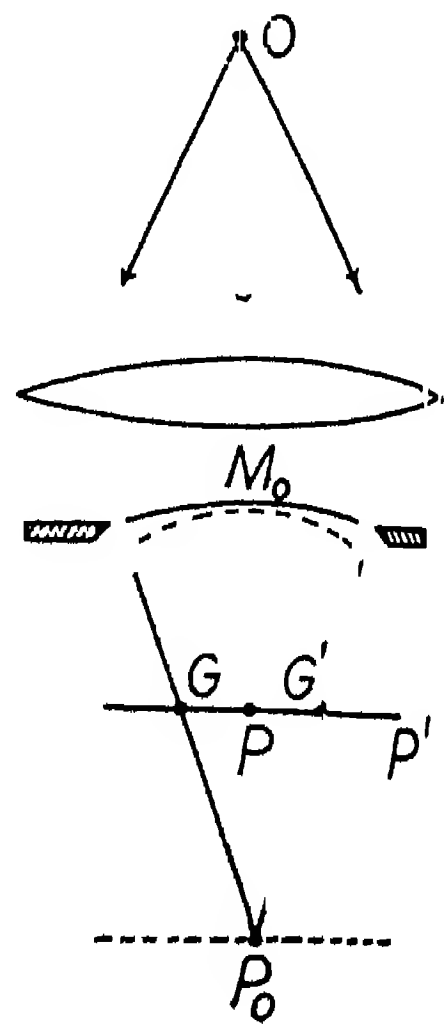


FIG. 99

We may obtain Fraunhofer patterns by using a single lens and placing the diffracting screen behind this (Fig. 97) or ahead of it (Fig. 98). In either case the actual wave front MN and the diffracted wave front ME are spheres having both the same radius of curvature. We may so to speak rotate the actual wave front about a hinge located at the left-hand edge of the aperture to obtain the diffracted wave front. The two fundamental conditions which we derived when the actual and the diffracted wave fronts are both plane still hold under these conditions, namely, the path difference, $\delta = QT$, is proportional to the distance $q = MQ$, measured across the wave front, and, also, all points lying upon an infinitesimal vertical strip on the actual wave front are at a common optical distance from the field point P. Consequently we may still apply the theory which was developed for plane

actual and diffracted wave fronts. We may apply the same theory fundamentally unaltered, and may hence properly classify the phenomena obtained under the conditions just described and those obtained when the actual and diffracted wave fronts are plane, all as Fraunhofer diffraction phenomena. Fraunhofer in his own observations used the arrangement of Figure 98. He placed the diffracting aperture before the objective lens of a telescope which had been previously focused on the source of light and made his measurements while viewing the pattern through the eyepiece of the telescope. We may use one lens, two lenses, or several lenses, placed much as we like with reference to the diffracting aperture—the pattern which is formed in the focal plane which is conjugate to the plane in which the source of light lies will be a Fraunhofer pattern.

The pattern which is observed in a plane which is *not* focally conjugate to the plane in which the source lies is of the Fresnel class: Referring to Figure 99, let O be the source and let P_0 lie in the conjugate focal plane. Let the pattern be observed in the plane PP', and let the diffracting aperture be a vertical slit. We may in this case apply the method developed in our study of the Fresnel pattern due to a vertical slit (chap. 4, sec. 2) by making suitable adaptations. The radius of curvature of the actual wave front, which we designated by a, must be given a negative sign because the center of curvature lies in *advance* of the wave, that is, $a=-M_0P_0$, and $b=+M_0P$. With this understanding we can make use of the Cornu spiral and in our calculations apply the equations by which the three quantities q, v, and z are interrelated. This simply means that the path increment is determined by equation (5) of chapter 2, namely:

$$\delta = q^2 \frac{(a+b)}{2ab} .$$

In order to apply the various equations in question, the term $(a+b)/a\,b$ must differ from zero. In other words, the pattern is of the Fresnel class when $(a+b)/a\,b \neq 0$, and this involves that $a \neq -b$, and that either $a \neq \infty$ or $b \neq \infty$. On the other hand, when the pattern is observed in a plane which is focally conjugate to the plane in which the source lies, then $(a+b)/a\,b=0$, the pattern is of the Fraunhofer type, and either $a=\infty$, and $b=\infty$, or $a=-b$.

Referring again to Figure 99, the edges of the geometrical shadow lie at G and G' in the Fresnel pattern obtained in the plane PP'. If this plane were slowly moved toward P_0, these edges would approach each other and in the position in which the Fraunhofer pattern is obtained they would coincide. Or, as has been previously pointed out, in Fraunhofer patterns, the geometrically illuminated region has collapsed to a single point (or to a vertical line when a vertical slit is used as the source).

Figure 99 represents only one set of various possible conditions for obtaining a Fresnel pattern when a lens is used. Other conditions may prevail when a single lens is used—or several lenses may be used.

A recapitulation of the various facts above set forth is given in Table IV.

TABLE IV

RECAPITULATION

Fraunhofer Diffraction	Fresnel Diffraction
1. A lens or a lens system is employed.	1. No lenses employed—a partial truth.
2. The actual and the diffracted wave fronts are either both plane, or have equal curvature, with the centers of curvature both lying on the same side of the diffracting screen.	2. The actual and the diffracted wave fronts have different curvature, in general. When the centers of curvature lie on opposite sides of the diffracting screen, they may, however, have equal curvature, as a special case.
3. Fraunhofer diffraction may be regarded as that special case of diffraction without lenses, Fresnel diffraction, in which both the source of light and the field plane lie at infinity.	
4. Fraunhofer diffraction is accommodated or focused, meaning that the source of light and the field plane lie in conjugate focal planes (and *not* meaning that the resulting fringes have sharp boundaries).	4. Fresnel diffraction is not accommodated—not focused, or out of focus.
5. $\delta = \text{Const.} \times q$	5. $\delta = \text{Const.} \times q^2$
6. $\frac{a+b}{ab} = 0$	6. $\frac{a+b}{ab} \neq 0$
Either $a = \infty$ and $b = \infty$ or $a = -b$	$a \neq \infty$ or $b \neq \infty$ and $a \neq -b$
7. The *geometrically* illuminated region collapses to a point, or a line.	7. The geometrically illuminated region is an *area* bounded by the "edge of the geometrical shadow."

8. Virtual Diffraction Patterns.—Suppose that an adjustable slit is held vertically before the eye, and while this slit is wide open the eye is focused upon a vertical lamp filament, O, perpendicular to the plane of the drawing (Fig. 100). If now the slit is narrowed, the Fraunhofer diffraction pattern due to a narrow rectangular aperture is formed upon the retina of the eye,

R. This is a *real* diffraction pattern. However, this pattern *appears* to the observer to be located in the plane *OQ* in which the source lies. No pattern actually exists in this plane—the pattern is *virtual* in so far as its existence there is concerned. The phenomena here involved are obvious. However, the phenomena involved in the formation of virtual diffraction patterns of the Fresnel class are not obvious. We shall now consider these.

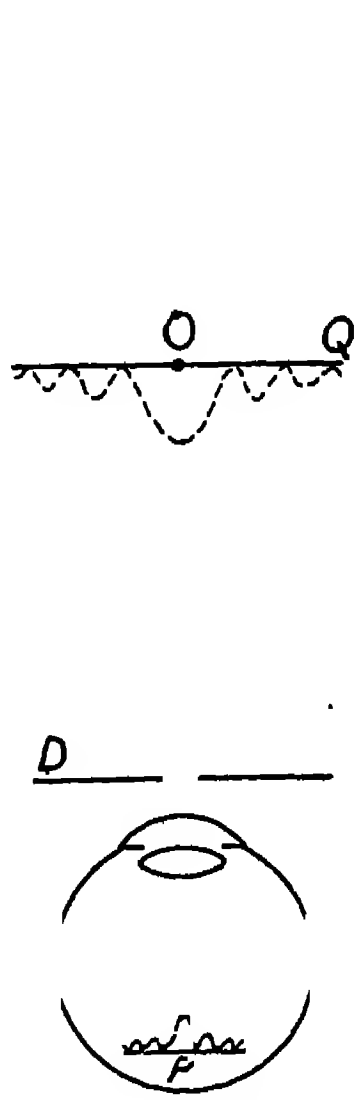


FIG. 100

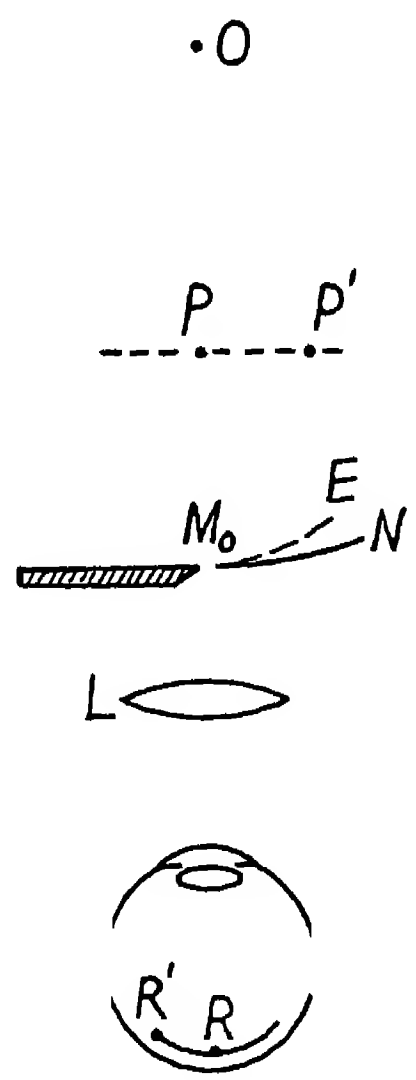


FIG. 101

In observing a *real* diffraction pattern of the Fresnel class—for example, that due to a straight edge—an eyepiece or hand lens is commonly mounted at a considerable distance from the straight edge and the pattern is viewed through this. For our present purpose we may imagine that we are using a reading lens of some four- or five-inch focal length. If the lens and eye together are advanced toward the straight edge, the fringes crowd together, and they disappear when the principal focal plane of the lens coincides with the straight edge. In this position the straight edge is seen in sharp focus. If the lens and eye are further advanced, very narrow fringes reappear, and these broaden as the advance is continued. Referring now to Figure 101, under the last-mentioned conditions the eye focuses through the lens on the field plane *PP'*. This plane and the retina of the eye, *R'R*, are focally conjugate with reference to the lens and eye considered as an optical unit. The pattern is seen as though it were located in the plane *PP'* where no pattern really exists. The pattern is virtual in so far as this plane is concerned. However, we know that there must be a real pattern formed on the retina.

We can determine the nature of this real pattern by an indirect method. Disturbances from various points on the actual wave front M_0N will reach a point on the retina—for example, point R—in different phases. The illumination at R will be determined by these phase differences. Further, since R is conjugate to P, all optical paths from P to R have the same length. Consequently, *if* the actual wave front M_0N could be made to propagate itself *backward*, the disturbances from the actual wave front would arrive at P with the same phase differences. The illumination which actually prevails at R would then prevail at P. In the same way, the illumination which prevails at R' would prevail at P' if the actual wave should travel backward. In other words, we may employ the Cornu spiral to determine what the pattern in the plane PP' would be if the wave should travel backward. In doing this we must take $a = -OM_0$ and $b = PM_0$. The pattern which would then be formed is identical with the virtual pattern which the eye perceives in the plane PP', and the real pattern which is formed on the retina is the ordinary optical image of this.

9. The Symmetry Properties of Fraunhofer Patterns.—A certain type of symmetry exists in Fraunhofer patterns even when a diffracting aperture is used which possesses no symmetry of any type. Consider the quasi-triangular aperture represented by the solid line contour of Figure 102 to be divided into strips of equal width as indicated. In the plane of the pattern

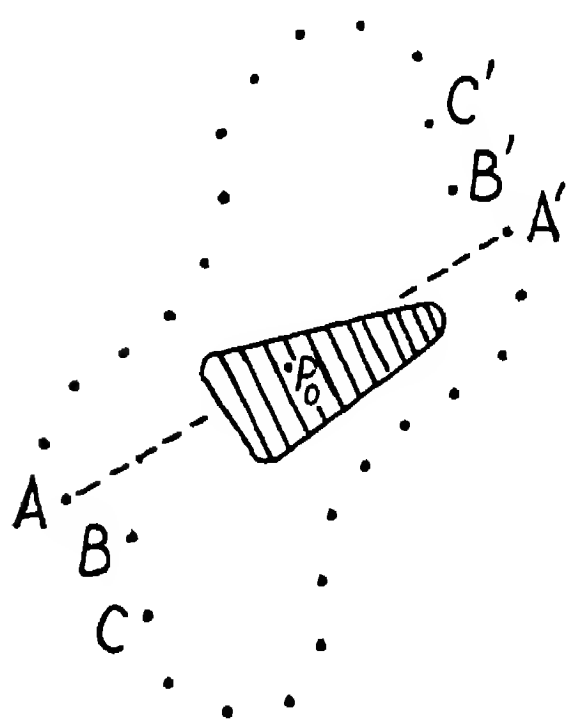


FIG. 102

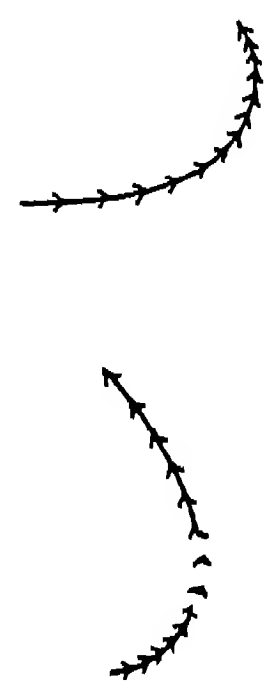


FIG. 103

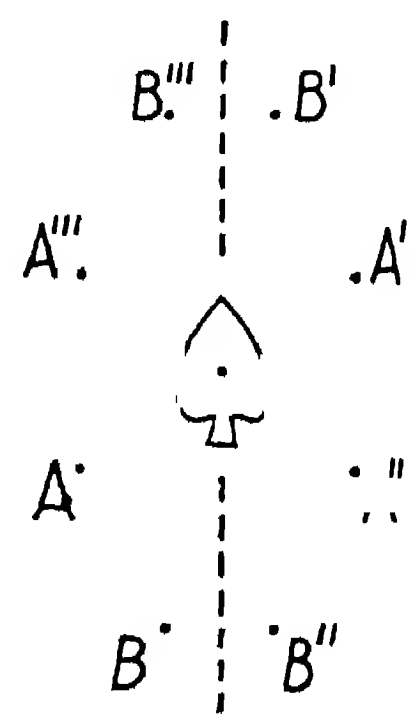


FIG. 104

erect a line through the central point P_0 of the field plane, perpendicular to the strips. Let us now inquire how the illumination compares at two points A and A' equidistant from P_0. If the vibration polygon for point A is as represented in the upper part of Figure 103, then the vibration polygon for point A' will be as represented in the lower part of the figure. The polygons have closing vectors of the same length and hence equal illumination prevails at points A and A'. Similarly, the illumination at point B' must be the same as at point B and that at C' must be the same as that at C. Thus,

point symmetry exists about the point P_0 even when the aperture is without symmetry.

When the aperture has a line of symmetry (see Fig. 104) we can draw a further conclusion. According to the principle just established, point A' in the pattern must have the same illumination as A. And because of the symmetry of the aperture A'' must have the same illumination as A, and A''' must have the same illumination as A'. Therefore all four points have equal illumination. In the same way points B, B', B'', and B''' must all have equal illumination. Therefore when the aperture has a single line of symmetry, the pattern has one line of symmetry parallel to this and another perpendicular to it.

A rectangular aperture has two lines of symmetry perpendicular to each other. Each of these gives rise to one line of symmetry in the pattern parallel to itself and another perpendicular to itself. The pattern has thus in a sense four lines of symmetry, but these coincide in pairs so that actually the pattern has only two lines of symmetry.

An aperture in the form of an equilateral triangle has three lines of symmetry each of which gives rise to two distinct lines of symmetry in the pattern (see Fig. 105); the symmetry of the pattern is thus sixfold.

FIG. 105.—The pattern due to an aperture having the form of an equilateral triangle (Scheiner and Hirayama).

When the aperture has n lines of symmetry, and n is an odd number, none of these lines will be perpendicular to each other and hence the pattern will have $2n$ distinct lines of symmetry. But when n is an even number, the lines of symmetry of the aperture will be perpendicular to each other in pairs, and hence the lines of symmetry of the pattern will coincide in pairs and the number of distinct lines of symmetry of the pattern will be only n.

The symmetry properties of Fraunhofer diffraction patterns were first discussed by R. Straubel.*

10. The Theorems of Bridge.—The English astronomer John Bridge† stated five general theorems relating to Fraunhofer diffraction patterns which are as follows:

* Dissertation (Jena, 1888) and *Ann. d. phys.*, **56**, 746, 1895. Interesting photographs are shown by Scheiner and Hirayma, *Abhand. d. Berliner Akad.* (1894), Anhang, p. 1. For further references see Geiger and Scheel, *Handb. d. Phys.*, **21**, 927.

† *Phil. Mag.* (4), **16**, 321, 1858.

1. Two apertures which have the same form but differ in size yield diffraction patterns which are similar but differ in size; the sizes of the patterns are in inverse proportion to the sizes of the apertures.

2. When the light is monochromatic, and various wave-lengths are used in succession, the patterns formed by a given aperture are similar, and the size of each pattern is in direct proportion to the wave-length.

3. The nature of the pattern formed by a given aperture does not depend upon the particular position of the aperture within the aperture of the optical system, or, when a given aperture is displaced parallel to itself across the aperture of the optical system, the pattern does not alter in any way.

4. When the value of every ordinate of the aperture is altered by a constant factor m, the amplitude at every point along the central horizontal line of the pattern is altered by the same factor m and hence the intensity at every point of this line is altered by m^2.

5. Given an opaque screen which has in it a number of apertures all having the same form and size and all similarly oriented, the intensity at any point of the focal plane may be considered as the product of the intensity which any one aperture acting alone would yield and the intensity which would be obtained from a point system having one point for each aperture and with the points situated at corresponding positions of adjacent apertures.

Proof of the foregoing theorems will not be given because the first four are practically self-evident from the principles of Fraunhofer diffraction as previously outlined, and the truth of the fifth theorem will become evident when an expression for the intensity yielded by a grating consisting of alternate transparent and opaque strips is derived, as will be done in section 15 of chapter 6. Detailed proof of all of the theorems may be found, however, in the reference given below.*

11. The Life and Work of Joseph von Fraunhofer.—Fraunhofer was a manufacturing optician. He was born in the town of Straubing, Bavaria, on March 6, 1787, and was the eleventh child of a poor glazier. The boy was of delicate physique and as a mere child was obliged to act as helper to his father. When he was about eleven years of age his parents died in short succession and thereupon those who were appointed as his guardians apprenticed him to a master-worker of cutglass in the city of Munich. Here, as was customary in the day for an apprentice, he lived with the family of his master. In July, 1801, the old house in which the family lived collapsed, burying the mistress of the house and the boy Joseph in the débris. The mistress died of her injuries but the boy suffered only minor lacerations. The elector Maximilian, who later became king of Bavaria, was present

* Verdet, *Leçons d'optique physique*, 1, 315.

when the boy was extricated from the ruins, as was also one J. Utzschneider who was a man of position and influence in the city of Munich. These men, taking an interest in young Fraunhofer, saw to it that he was subsequently allowed to attend night and holiday schools. His schooling had up to that time been extremely meager. The elector Maximilian presented the boy with some gold of which he later expended a portion to buy himself free from the last half-year of his apprenticeship and expended the remainder for the purchase of a machine for grinding and polishing spectacle lenses.

Completing his apprenticeship in May, 1804, when he was seventeen years of age, young Fraunhofer set out, in times which were bad because of the Napoleonic wars, to earn his own living as a journeyman worker in cut-glass and by engraving. He eked out the barest existence at this for two years and devoted his spare time to study. His talents were recognized by those who had come to know him, and in May, 1806, the aforementioned Utzschneider offered to employ him as an optician in the mechanical and optical manufacturing firm of Reichenbach, Utzschneider and Liebherr, which had been founded a few years previously. Fraunhofer accepted this employment. In the following year the optical manufacture of the firm was removed to Benediktbeuren, where the glass furnaces were already located. This village was at some distance from Munich and Fraunhofer moved thither. Here he was placed in a position of increased authority and one which permitted him some independence of action. In the year 1819 the optical manufacture was moved back to Munich and Fraunhofer moved back with it. The genius of Fraunhofer made itself felt from the outset of his employment by the firm and resulted in a great enhancement in the quality of the optical products, with a consequent rapid rise in the demand for these. Fraunhofer was soon placed in full technical charge and, in 1814, after a reorganization, became a partner of the optical manufacturing firm of Utzschneider and Fraunhofer.

Fraunhofer is famous both for the improvement which he wrought in the astronomical telescope and for several advances in pure science. His advances in pure science were made in close association with his efforts to improve the telescope, but these advances were not regarded by him merely as a means to an end. On the contrary, he, toward the end of his career, cherished longings to be free from his work of optical manufacture and devote himself to scientific pursuit, but this longing was never realized.

The first task of importance performed by Fraunhofer was to determine the index of refraction for various colors of each of given types of optical glass. In the course of this task he observed in the continuous spectrum of a lamp a line in the orange yellow which was brighter than the remaining spectrum. Further observation revealed that this same line was present in the spectrum of *various* types of flames. The line was what we today know

as the sodium line, and Fraunhofer's observation constituted the discovery of this important spectrum line. He then used the line as a reference mark and, by means of a crude but ingenious piece of apparatus, was able to determine the indices of refraction at various well-established positions of the spectrum, thus overcoming a difficulty which had theretofore always existed, namely, the difficulty of specifying precisely the spectral position to which a given index determination applies. Subsequently Fraunhofer proceeded to use sunlight, wishing to ascertain whether the same orange-yellow line could be found in sunlight. To his surprise he found a dark line at the same position where the bright line appears in the flame, and in addition he found numerous other dark lines at other positions. These dark lines are what we now call the Fraunhofer lines. In all subsequent work these dark lines served him as reference marks in the spectrum. Fraunhofer's account of the discovery of these lines was presented in 1817; the discovery itself had been made some two or three years earlier.*

Fraunhofer was the first to combine a narrow slit, a prism, and a telescope in such a manner as to form what we would today call a "spectroscope." It may thus be claimed that he invented the spectroscope—but this claim must be advanced with some reserve. Newton, long before Fraunhofer, had used a slit, a prism, a lens, and a diffusing screen in combination, in a way which may also be claimed to constitute a very crude spectroscope. The development of the spectroscope was to a large extent a process of gradual evolution.

A further achievement of Fraunhofer is that he succeeded in greatly improving the quality of optical glass. He improved the processes of melting and of casting glass with the result of procuring much larger masses free from striae. His improvement of optical glass and his development of methods of computing accurately the forms which the elements of the lens should have and of methods of accurately figuring each element according to the computed form, brought the astronomical telescope to a perfection and size theretofore undreamed of. Before the day of Fraunhofer a diameter of about five inches was the maximum, whereas Fraunhofer raised the maximum to nearly ten inches, at the same time reducing the aberrations.

Fraunhofer, like Fresnel, undertook experiments on diffraction with a view to learning more of the nature of light through these experiments. His general objective was to make advances in the study of diffraction by applying optical instruments to this study. Before Fraunhofer, diffraction phenomena had been studied without optical instruments, or, at most, by the aid of an eyepiece, following the method which had been introduced by

* Before Fraunhofer, the London physician, Dr. W. H. Wollaston, had discovered seven dark lines in the solar spectrum—in 1802. Wollaston's discovery, though published, was not followed up, and was forgotten until Fraunhofer rediscovered the lines.

Fresnel in 1815 (as described in sec. 4 of chap. 1). The results of Fraunhofer's investigations on diffraction were submitted in the years 1821 and 1822.

Fraunhofer, in studying the pattern due to a slit, sought to obtain this pattern with greater intensity. In order to achieve the desired result he made a series of slits close together by winding a hair or wire upon a frame. What he then found was not at all what he had expected. The pattern was *not*, as he expected, that due to a single slit only more intense; it was that due to a *diffraction grating*—he had discovered the diffraction grating. His discovery of the grating was thus to a large extent accidental, but he showed great genius in the manner in which he followed up this discovery as well as in the manner in which he followed up his discoveries of the sodium line and of the dark lines in the solar spectrum which were also to a certain degree accidental.*

Fraunhofer, like Fresnel, contracted tuberculosis, and died at the same age as Fresnel, namely, at thirty-nine years. He died on June 7, 1826. Like Fresnel also he was never married. He was one year older than Fresnel. Shortly before his death he was made a knight.

The works of Fraunhofer were collected and published by the Munich Academy of Science in 1888 under the title *Joseph von Fraunhofers Gesammelte Schriften*. This volume also contains a biography. There are numerous other accounts of the life and work of Fraunhofer of which the most recent and comprehensive is a book by M. von Rohr (Leipzig: Akad. Verlagsgesell., 1929).

In concluding the present chapter be it mentioned that a book devoted to Fraunhofer diffraction phenomena, entitled *Die Beugungserscheinungen*, was written by F. M. Schwerd and published at Mannheim as early as the year 1835.

* The grating had been previously discovered by the American astronomer, David Rittenhouse, but Rittenhouse's discovery failed to attract any attention (in this connection see chap. 6, sec. 10).

CHAPTER 6

THE DIFFRACTION GRATING

The diffraction grating may take any one of a large number of forms. In its most prized form it is produced by ruling a multitude of fine, parallel, equidistant lines, by means of a diamond point, upon a plate of polished spectulum metal. The operation is performed by a "ruling machine." The grating serves to produce a spectrum and is by far our most valuable instrument of spectroscopic investigation.

Gratings ruled upon speculum receive the incident light and diffract it by *reflection*. Some types of gratings diffract the incident light by *transmission*, such as, for example, the gratings commonly used in student laboratories. These are copies or "replicas" made from original, ruled gratings. The method of making replicas will be described later. Another form of transmission grating may be made by depositing a film of silver upon a plate of glass and ruling lines through the film. The film then constitutes an opaque screen having narrow slits in it which are of equal width and are separated by small equal distances. The light is diffracted by the slits. The theory of the grating can be most lucidly presented by considering first a grating of this form—by explaining the diffraction effects produced by the slits.

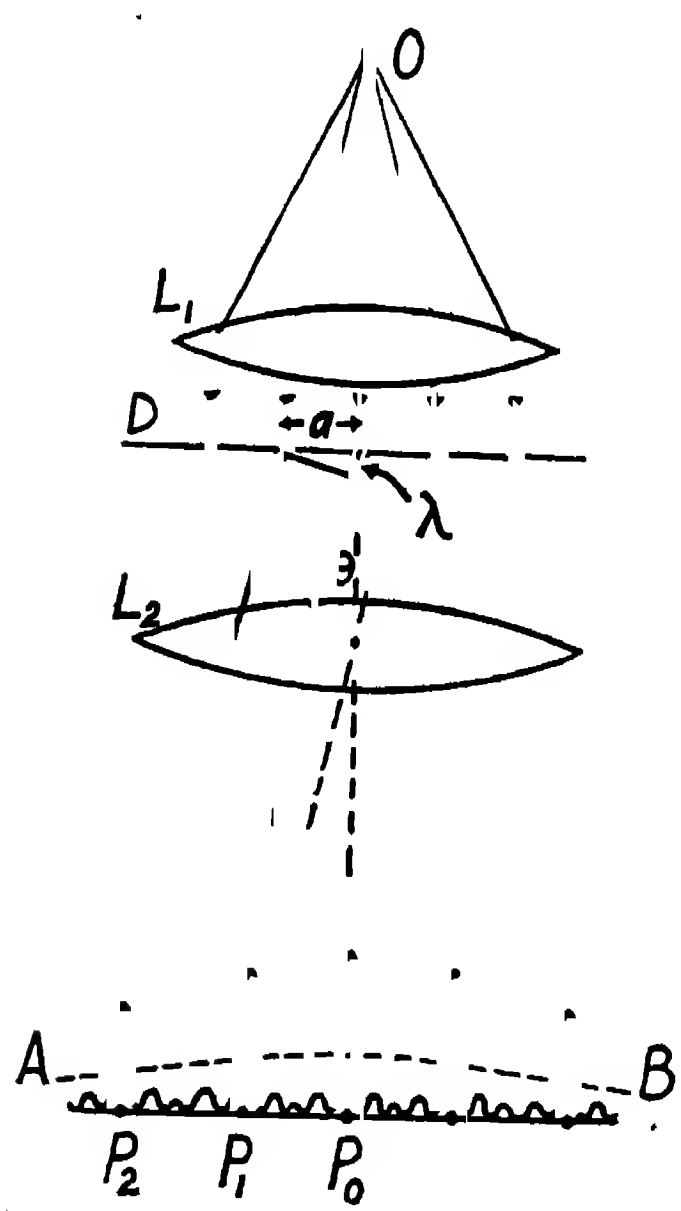


FIG. 106

1. Diffraction by Narrow Equidistant Slits.—Referring to Figure 106, let us suppose that O is a linear vertical source of light and that the light, after being made parallel by the lens L_1, is incident perpendicularly upon the opaque diffracting screen D which contains the fine slits or the "lines" of the grating. Beyond the screen, parallel diffracted rays are focused by the lens L_2 in the plane of observation $P_0P_1P_2$. The diffraction patterns formed by a grating fall into the *Fraunhofer class* of diffraction phenomena. The fringes run perpendicular to the plane of our figure, or vertical. We shall

suppose for the present, first, that the light is monochromatic, there being then no "spectrum" formed; and, second, that the slits are so narrow that each single one approaches the condition of sending Huygens' wavelets in all lateral directions. The distribution of light from one slit acting alone will then be somewhat as represented by the height of the broken curve AB.

A wave front reaching the diffracting screen gives rise to disturbances which issue from the various slits simultaneously. These disturbances will reinforce each other at the point P_0, which lies at a common optical distance from all of the slits. The disturbances will also reinforce each other at certain other points, P_1, P_2, etc., so located that the difference in optical path from adjacent slits is respectively λ, 2λ, etc., or in general, $m\lambda$, where m is an integer which is known as the *order* of diffraction. The distance between corresponding points of adjacent slits, let us say between centers, is known as the "grating space." We shall denote this by the letter a, and denote the angle of diffraction by θ. For the maximum P_1 we now have $\lambda = a \sin \theta$, and for the maximum of mth order, for which the difference in optical path by way of adjacent lines is $m\lambda$, we have:

$$m\lambda = a \sin \theta . \tag{1}$$

This is the form which the so-called "grating law" takes when the light is incident perpendicularly. This law determines the *directions* in which the principal maxima are formed. The positions of these maxima in the field plane, e.g., the distance P_0P_1, can be readily ascertained when the focal length of the lens L_2 is given. It is to be noted that the directions and positions of the principal maxima do not depend upon the number of lines in the grating and that they do depend upon the wave-length of the light.

2. Formation of Spectrum.—The number of lines in the grating enters as an important factor in determining the complete form of the diffraction pattern, involving, for example, the *breadth* of the principal maxima and the number and relative height of the subsidiary maxima which are represented in Figure 106 as occurring between the principal maxima. The explanation of the complete form of the pattern will be given later. It will develop that when the number of lines of the grating is large the principal maxima are very narrow and constitute sharp *images* of the linear source of light and that the subsidiary maxima then ordinarily escape observation. The linear source is usually the slit of a spectroscope and hence the principal maxima are often spoken of as "slit images." The maximum of zero order is known as the "central image" or "undeviated image." The maximum P_1 is the "first-order image," the maximum P_2 is the "second-order image," etc.

When the source of light is not monochromatic, a slit image is formed in each order by every wave-length which is present in the source. A spectrum

is formed. When the wave-lengths present in the source are discrete, that is, separated, the slit images are separated and constitute "spectrum lines." When the wave-lengths cover a certain range continuously, the slit images overlap—a continuous spectrum is formed. At the top of Figure 107 are shown the slit images P_0, P_1, etc., as they would be seen in the plane of observation when the light is monochromatic. Below these are shown the slit images or spectrum lines which occur when there are three discrete

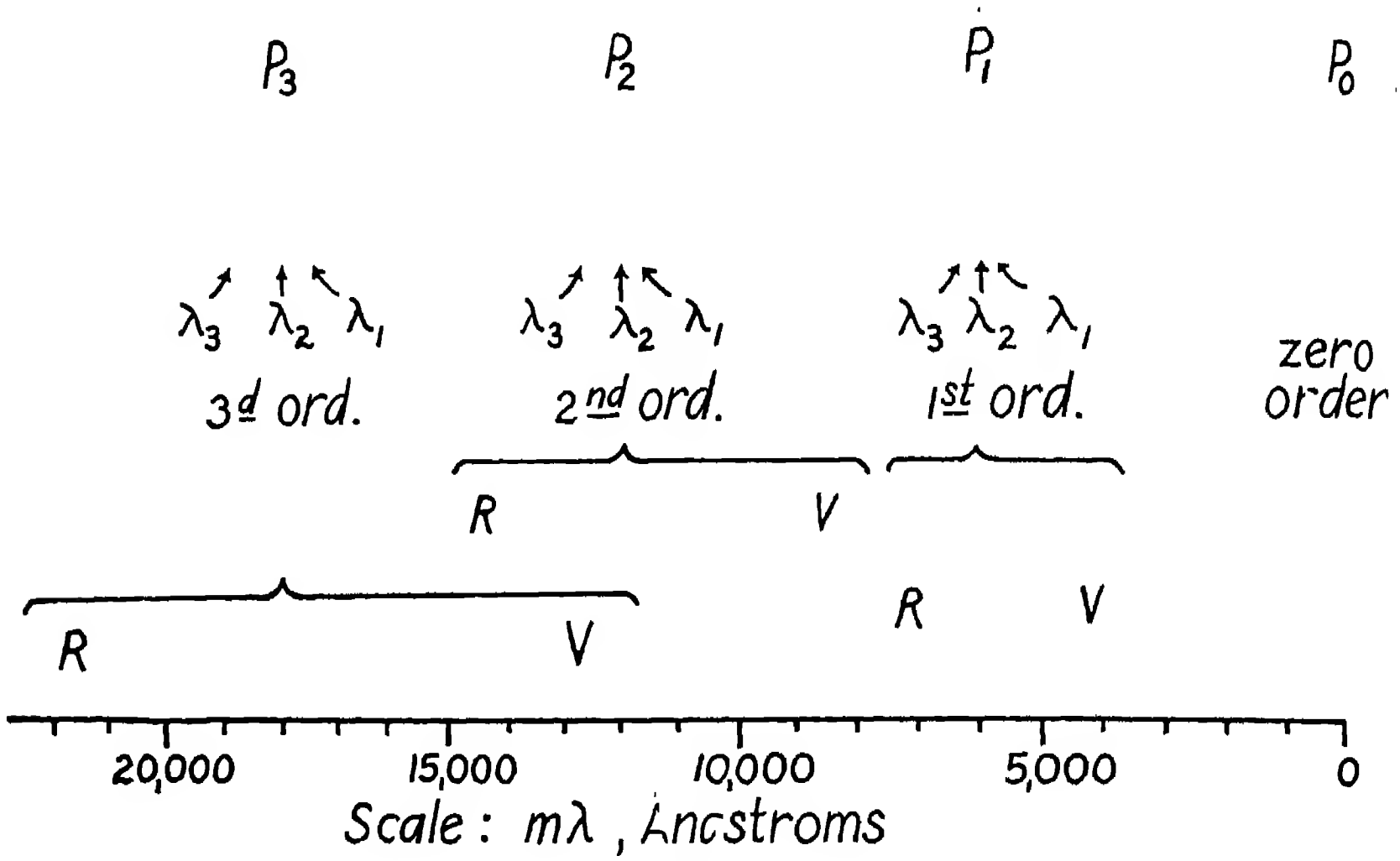


FIG. 107

wave-lengths in the source which do not differ greatly from each other. The wave-lengths are denoted by λ_1, λ_2, and λ_3, and supposedly $\lambda_1 < \lambda_2 < \lambda_3$. In the zero order the three images are superposed in a single "central image." In the other orders the images are separated, and, as follows from equation (1), they are separated by distances which are very nearly proportional to the difference between the wave-lengths in question. Moreover, it follows that in the second order the distance between two images will be approximately twice as great as in the first order, and in the third order the distance will be approximately three times as great, etc. The *dispersion* due to a grating increases in proportion to the order of the spectrum.

If the emission lines are distributed throughout the visible range there will be a complete spectrum formed in each order as illustrated in the lower part of the figure. The region of short wave-lengths in each spectrum is

marked V, for violet, and the region of long wave-lengths is marked R, for red.

3. Overlapping Spectra.—In the lower part of Figure 107 the third-order spectrum has been drawn below the other orders because the violet end of this spectrum overlaps with the red end of the second-order spectrum. This overlapping or superposing of orders is a general feature of grating spectra. The first and second orders do not overlap in the visible region. The spectrum physically considered, however, extends into the ultra-violet and infra-red, and consequently when photographic or bolometric methods of observation are employed, overlapping between the first and second orders must also be taken into account. For example, the first order of wave-length 7400 Å, in the deep red, will coincide in position with the second order of half this wave-length, namely, 3700 Å, which lies in the ultra-violet only a little beyond the visible limit. Therefore, in photographing in the red region, care must be taken to absorb all of the ultra-violet by a filter.

In general, a spectrum line of wave-length λ, in the spectrum of the mth order, coincides in position with a line of wave-length λ' in the order m' when the condition is satisfied that $m\lambda = m'\lambda'$.

The so-called "method of coincidence" was used by Rowland for determining the wave-lengths of spectrum lines with accuracy far exceeding previous determinations. Rowland's work extended through the decade 1886–96, and led to the establishment of the "Rowland scale" of wave-lengths. In applying the method of coincidence it is not necessary to find spectrum lines which actually coincide in different orders—they need only fall near each other. Interpolation can then be applied. The method of coincidence has now been superseded by interference methods for establishing wave-length *standards* in those spectral regions for which the photographic plate is sensitive or can be sensitized, but the method of coincidence is still much used in the routine of determining wave-lengths in an unknown spectrum by comparison with a known spectrum. Moreover, in the infra-red beyond 12,000 Å (1.2 μ) where photography fails, the method of coincidence is still the best available.

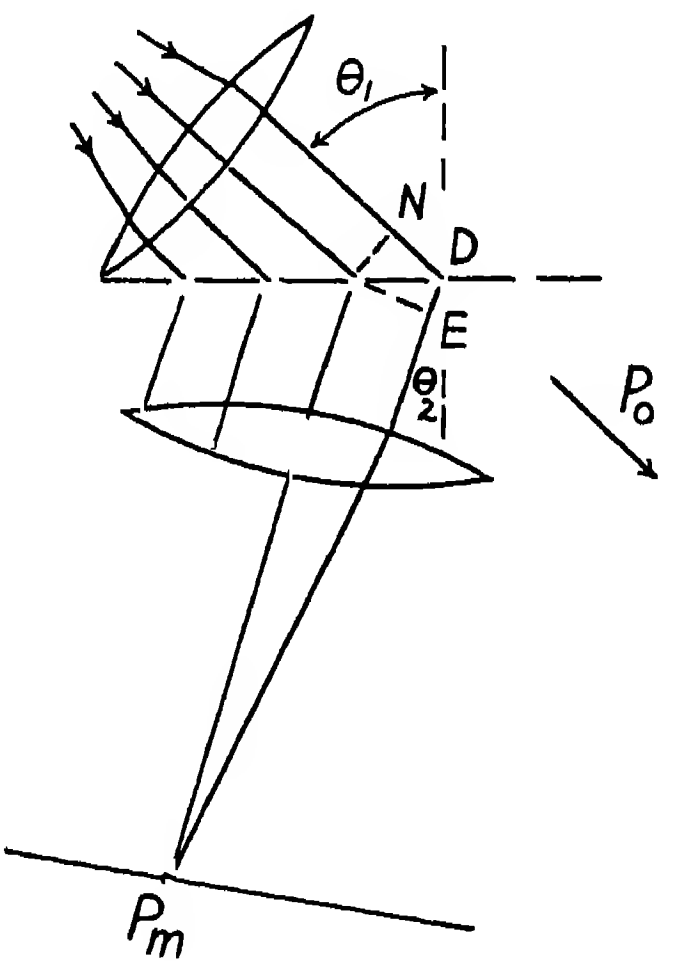


FIG. 108

4. The Grating Law.—Gratings are not generally used with the light incident perpendicularly, as we have thus far supposed. The light may be incident at any angle, θ_1 (see Fig. 108), and diffracted at any angle θ_2. The path difference by way of corresponding points of adjacent lines is then $ND+DE$. The principal maximum of the mth order of the wave-

length λ will be formed when $ND+DE=m\lambda$. Now $ND=a \sin \theta_1$ and $DE=a \sin \theta_2$, where a is the grating space. We therefore have:

$$m\lambda=a(\sin \theta_1+\sin \theta_2) . \tag{2}$$

This equation is known as the "grating law." When the incident beam and the diffracted beam both lie on the same side of the normal to the grating, as illustrated, both angles are to be taken as positive. When the beams lie on opposite sides of the normal, the smaller of the two angles is to be taken as negative.

When the angles θ_1 and θ_2 both approach 90°, the path difference by way of adjacent lines of the grating approaches the value $2a$. If m equals unity, the value of λ approaches $2a$. The wave-length $\lambda=2a$ is the greatest wave-length which a grating can diffract.

The treatment of the grating considered as a diffracting screen will now be resumed. It will be supposed that the light is monochromatic, and the manner in which the diffraction pattern depends upon the number of lines of the grating will be investigated.

5. Gratings Having a Small Number of Lines.—We shall consider in succession gratings having 2 lines, 3 lines, 4 lines, 5 lines, and 6 lines.

The 2-line grating.—The Fraunhofer pattern for a double slit has already been treated.* In discussing this we did *not* assume that the slits are very narrow and close together. The pattern then considered is again shown in Figure 109*a*. The broken curve AB represents the distribution of illumination from each slit acting alone. This shows well-marked maxima and minima. The solid curve P_0, P_1, etc., represents the closely spaced maxima which arise from interference of the disturbances from the two slits. When the slits are greatly narrowed, the central maximum of the broken curve spreads laterally over the entire field (see Fig. 109*b*). The spacing of the maxima of the solid curve is, however, the same as when the slits are wide, provided the distance between centers of the slits is the same in both cases.

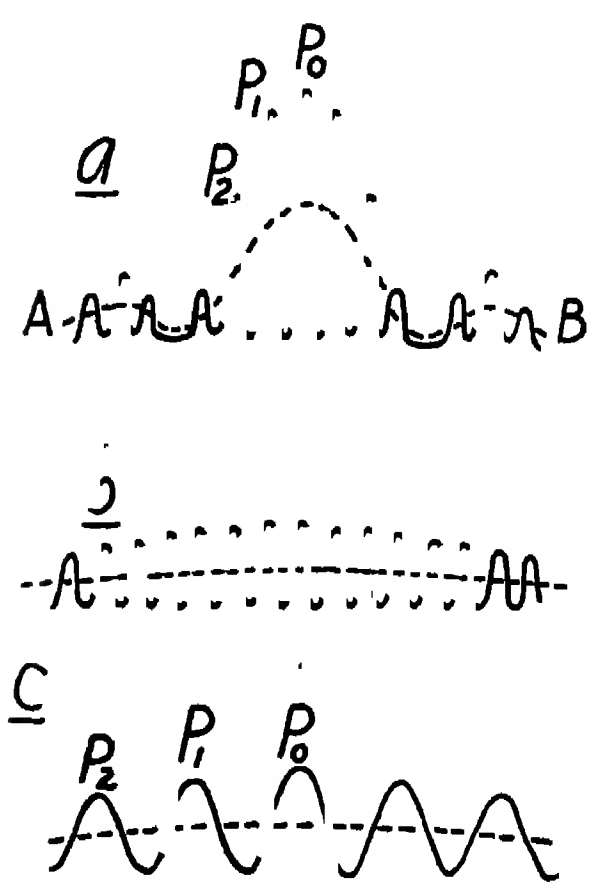


FIG. 109

If now the slit *separation* is decreased, then the separation of the maxima of the solid curve increases; see Figure 109*c*, which represents the pattern due to a 2-line grating under the conditions which we now assume, namely, that

* Chap. 5, sec. 5.

the slits are very narrow and close together. The left-hand portion of this curve is plotted to a larger scale in Figure 110. Here, also, at several points of the curve, the vibration diagrams which obtain at these points are inserted. Each vibration diagram consists of two vectors which represent respectively the disturbances from the two slits.* At the central point of the pattern P_0 the disturbances from the two slits arrive in the same phase. Thus both arrows point in the same direction. The vector which represents the disturbance from the left-hand slit is throughout plotted pointing horizontally to the right. Upon moving from the central point of the pattern to the left, the phase of the disturbance from the right-hand slit falls behind—the second arrow rotates counter-clockwise. Accordingly, as we go down the slope of the central maximum, the second arrow is shown successively as making an angle of 45° with the first, then 90°, then 135°. At the first minimum it makes an angle of 180°, or points horizontally to the left. The sum of the two vectors represents the resultant amplitude of illumination. This is now zero. Beyond the first minimum, further stages of rotation of the second vector are shown. When it again points horizontally to the right we have arrived at the first lateral maximum in the pattern, P_1. Beyond P_1 the cycle of rotation is repeated. Throughout, the lengths of the vectors are supposedly proportional to the height of the broken line.

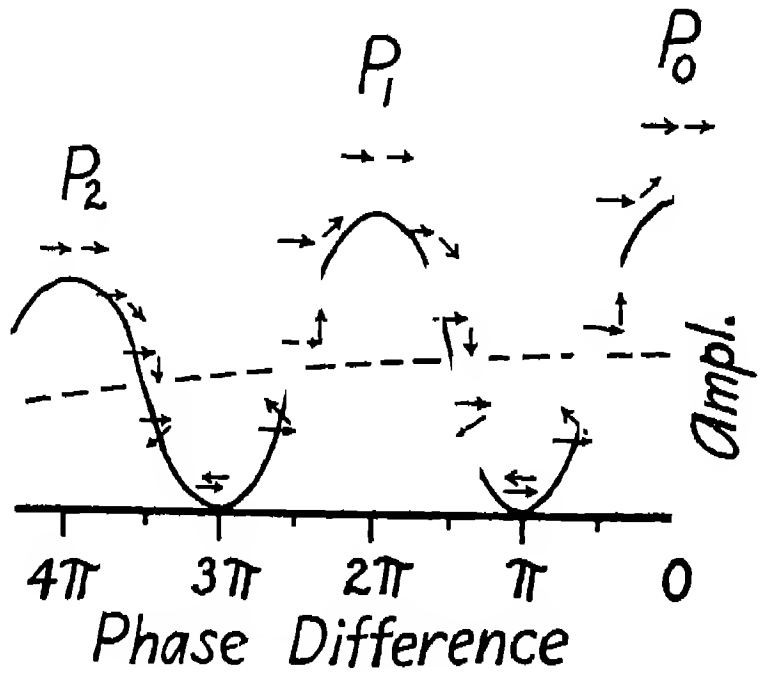


FIG. 110

The 3-line grating.—We now have three vectors, one for each line (see Fig. 111, upper left diagram). The vector corresponding to the left-hand line is again always drawn pointing horizontally to the right. At the central point of the pattern—right-hand end of the curve—the second and third vectors also point horizontally to the right. As we descend the slope of the central maximum we must think of the second and third vectors as rotating counter-clockwise, the third one rotating at double the rate of the second one, and the three forming at all stages an equiangular and equilateral polygon which is an open polygon while we are on the slope. Finally, at the first minimum this polygon closes, forming a triangle as illustrated. Beyond the first minimum, the counter-clockwise rotation of the second and third vec-

* We may picture each vector as the closing vector of a vibration curve drawn for the corresponding slit, in accordance with deductions in chap. 5, sec. 5. However, now that we are concerned with very narrow slits, it is simpler to consider the contribution from each slit as an entity unto itself represented simply by a vector.

tors continues and the sum subsequently attains a maximum value when the angle between successive vectors attains the value of π. The polygon has now collapsed to one in which the three vectors double back and forth upon each other (in the figure they are drawn above each other). The effects of the first and second vectors annul each other, leaving the effect of the third one outstanding. At this subsidiary maximum the amplitude of illumination is one-third of the amplitude at the principal maximum—the intensity is one-ninth of that at the principal maximum.

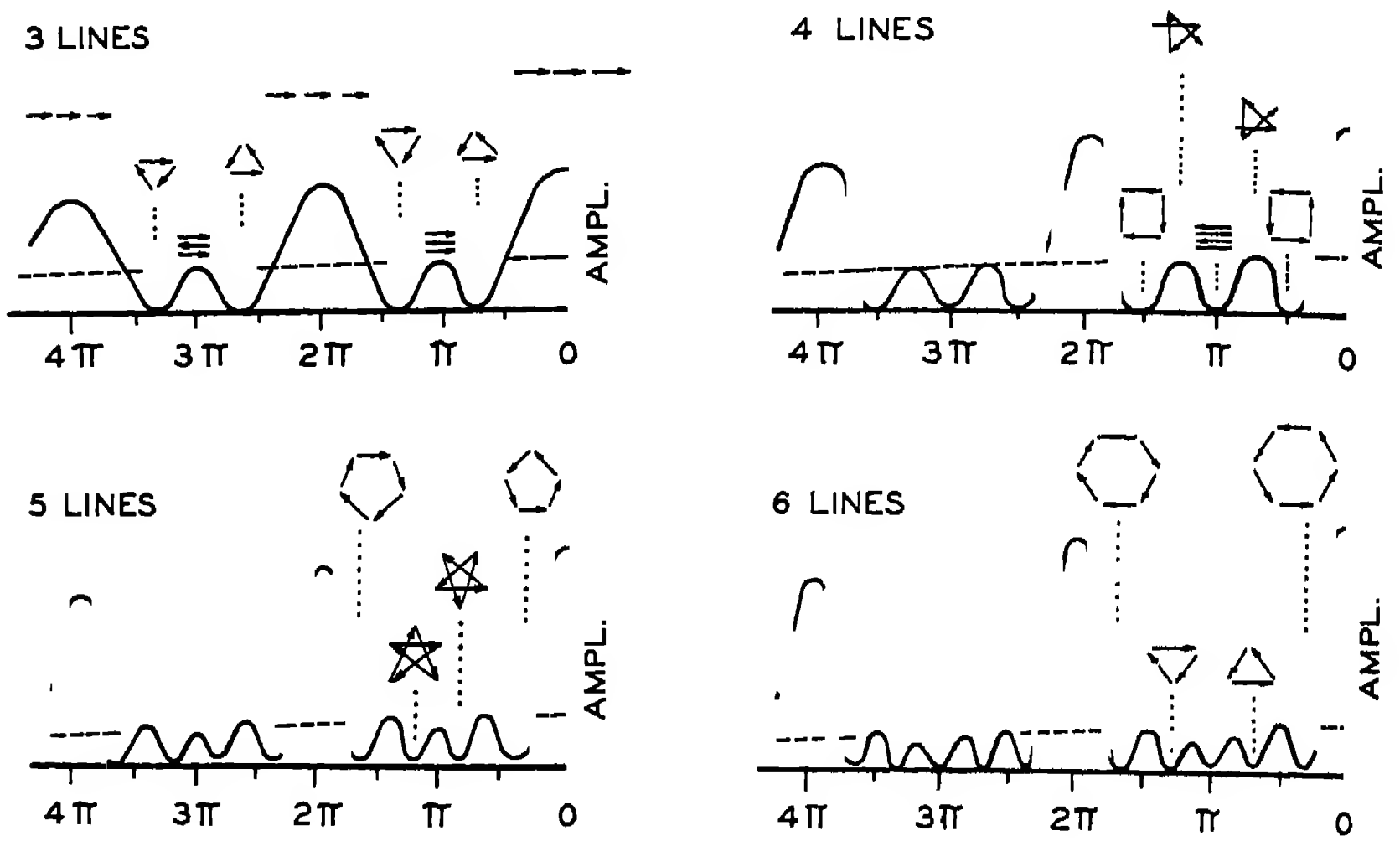


FIG. 111

Beyond this subsidiary maximum the counter-clockwise rotation of the second and third vectors continues, and when it has proceeded so far that an equilateral triangle is again formed, as illustrated, we are at the second minimum in the pattern. It is to be noted that this second triangle is described in the reverse sense from the first one, or clockwise. Beyond this second minimum, as the counter-clockwise rotation of the second and third vectors continues, the polygon opens up. We proceed up the slope of the principal maximum of first order, and when we reach this maximum the vectors again all point horizontally to the right. Beyond this point the cycle of rotation and the pattern are both repeated except for a modulating factor which is proportional to the height of the broken line.

The 4-line grating.—See upper right-hand corner of Figure 111. We now have four vectors. We reach the first minimum in the pattern when these vectors first form a closed polygon or, in other words, a square, as illustrated. Beyond this we reach a subsidiary maximum when the angle be-

tween successive vectors is $3\pi/4$, or 135° as shown. That this is true follows from the fact that the polygon, which is in general an open one, is at all stages both equilateral and equiangular and may hence at all stages be inscribed in a circle. The maximum occurs when the foot of the first vector and the head of the last vector lie at opposite ends of a diameter of the circumscribed circle (see Fig. 111*a*). Here this diameter is represented by the broken arrow S, which is the sum of the four constituent vectors. It is a simple matter to evaluate this sum and thus determine the height of the subsidiary maximum.* Beyond this subsidiary maximum the second minimum occurs when the angle between successive vectors is equal to π. The vectors now double back and forth upon each other as indicated in Figure 111. Beyond this minimum the curve is symmetrical with the portion which we have just considered. There are now two subsidiary maxima between the principal ones. The abscissa π is a point of symmetry in all of the curves which we are considering. There is a minimum at this point when the number of lines is even as now, and a subsidiary maximum when the number is odd.

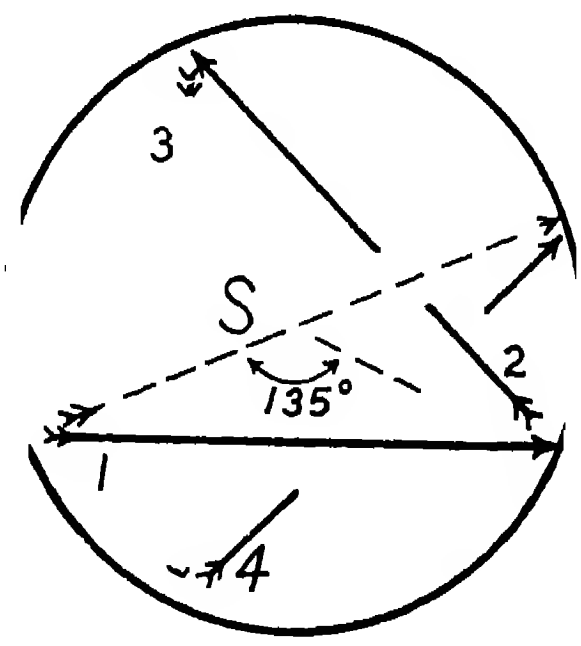


FIG. 111*a*

The 5-line grating.—The first minimum occurs when the five vectors in question form a pentagon as indicated at the lower left in Figure 111. The first subsidiary maximum occurs when the vectors form an open polygon of one and one-half convolutions. This polygon is not shown—the polygons at the maxima will henceforth be omitted. The second minimum occurs when the vectors form the next closed figure, a five-pointed star as illustrated. The next maximum occurs when the angle between successive vectors is π the vectors double back and forth on each other—the effect of one is left outstanding. Beyond this point the curve is symmetrical. There are now three subsidiary maxima.

The 6-line grating.—At the first minimum we have a hexagon as shown at the right. At the first subsidiary maximum the vectors form a square of one and one-half convolutions, and at the second minimum they form a triangle as shown, or rather two superposed triangles. At the next subsidiary maximum the polygon has two and a half convolutions and at the next minimum the angle between successive vectors is π—the vectors double back and

* Details follow: Denoting the length of each constituent vector by S_1, we have $(S_1/2)/(S/2) = \sin(135°/2) = .924$, whence $S = 1.08\,S_1$. The ratio of the height of this maximum to that of the central maximum is therefore 1.08/4 and the ratio of intensities is equal to the square of this, or .073.

forth on each other—their sum is zero. There are now four subsidiary maxima.

Each time another line is added to the grating there is an additional subsidiary maximum. When the number of lines in the grating is n, there are $n-2$ subsidiary maxima.

As we pass from the 3-line grating successively to the 4-, 5-, and 6-line gratings, there is a decrease of the intensity of the subsidiary maxima relative to the principal maxima. However, this decrease approaches a limit as the number of lines of the grating becomes very large—this fact will appear in detail later.

Finally, it is important to note that as the number of lines of the grating increases, the principal maxima become proportionately narrower. For a 3-line grating the first minimum in the pattern occurs when the phase difference between disturbances coming from adjacent lines is $2\pi/3$, or when the path difference is $\lambda/3$, for a 4-line grating, when the path difference is $\lambda/4$, and for a grating of n-lines, when the path difference is λ/n. The principal maxima thus become very narrow when the number of lines becomes large.

6. Gratings Having Many Lines.—Gratings used for spectroscopic purposes have many thousand lines—a large grating may have one hundred thousand. The resulting diffraction pattern cannot well be directly represented, but the principles involved in the formation of the pattern may be explained by considering a grating having a moderate number of lines. At the top of Figure 112 the pattern for a grating having twenty-four lines is shown. The central maximum, P_0, lies at the right. The first minimum is marked 1, the first subsidiary maximum, 1′, and the second minimum 2. In the lower right-hand corner of the figure the portion of the curve which includes these points is drawn to a larger scale and above the points are shown the corresponding vibration polygons, which have respectively one, one and a half, and two convolutions. In the upper portion of the figure the polygons which obtain at the fourth, sixth, and eighth minima are shown. These consist respectively of a hexagon having four convolutions, a square having six convolutions, and a triangle having eight convolutions. The convolutions are superposed vector for vector. At other minima the convolutions are likewise superposed but the individual vectors are not always superposed. The twelfth minimum is a point of symmetry in the pattern. Beyond it the vibration polygon uncurls progressively. When the polygon has completely uncurled, the vectors form a straight line, and this is the condition for the formation of the first principal maximum P_1. Beyond P_1 the cycle is repeated.

Let us now suppose that the number of lines of the grating is really large and see what conclusions we can draw.

The number of vectors of the vibration polygon is equal to the number of lines of the grating. At points in the pattern lying near a principal maximum—for example, at the points 1, 1′, and 2 of Figure 112—the number of convolutions of the polygon is small. Therefore, when the number of lines

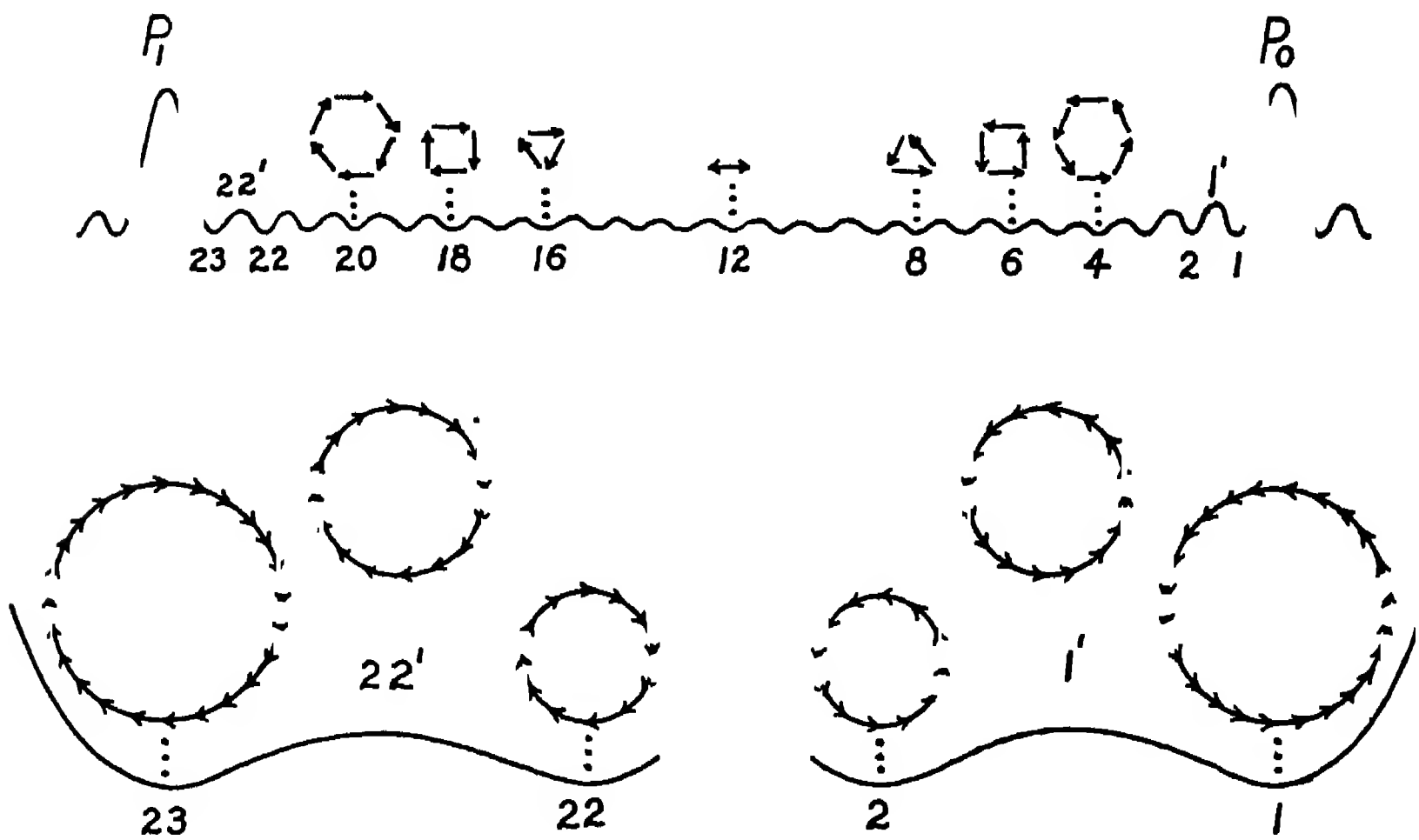


FIG. 112

of the grating is large, there will be many vectors per convolution and we may consider the vibration polygons for such points as vibration *curves*—of circular form. It will be recalled that the vibration curves for a rectangular aperture are likewise circular in form.* For both the grating and the rectangular aperture the first minimum occurs when the vibration curve closes in one convolution, the first subsidiary maximum occurs when the curve forms one and one-half convolutions, etc. The positions of the minima and both the positions and relative intensities of the subsidiary maxima will therefore be the same for the grating as for a rectangular aperture having a width equal to the width of the beam diffracted by the entire grating in the direction in question. This applies to the pattern in the vicinity of each of the principal maxima formed by the grating.

The beam diffracted by a grating is commonly some 5 or 10 cm. wide, and a rectangular aperture having the same width would yield an image of the source having a high degree of perfection. Accordingly, each of the principal maxima formed by a grating is a highly perfect image—a slit image.

* Chap. 5, sec. 2, Case A.

For a grating, as well as for a rectangular aperture, the intensity of the first subsidiary maximum relative to that of the principal maximum is $(2/3\pi)^2 = 1/22.2$ (see chap. 5, sec. 2 η). For the second subsidiary maximum the relative intensity is $(2/5\pi)^2 = 1/61.7$. The intensities of the subsidiary maxima of low order are great enough so that these maxima can be seen when they are separated from the principal maxima. In practice they are not usually separated. Thus with a grating they ordinarily escape observation just as they do when a wide rectangular aperture is used.

The subsidiary maxima become progressively fainter as their order number increases. The intensities and positions will be practically the same maximum for maximum as for the equivalent rectangular aperture while the order number is, let us say, below one hundred—there will still be many vectors per convolution of the polygon. The maxima in question are too faint to detect. As we proceed to still higher order numbers the intensities continue to diminish until we reach the point midway between two principal maxima. However, for subsidiary maxima of very high order we can no longer consider the vibration polygon even approximately as a curve—the polygon eventually has so many convolutions that there are only a few vectors per convolution. We can form an idea of the magnitude of the intensities which then prevail by evaluating the intensity of the subsidiary maximum which lies exactly midway between two principal maxima when we have a grating having an *odd* number of lines. At this point the vectors of the polygon double back and forth on each other—the effects of the lines cancel each other except for the effect of one line which is left outstanding. The relative intensity of the maximum in question is therefore $1/n^2$, or, if the grating has 100,001 lines, the relative intensity is about $1/10^{10}$.

Recapitulating: When a grating has many lines the principal maxima of diffraction are nearly perfect images of the source—they are slit images in monochromatic light. The subsidiary maxima are faint and cause no trouble. Those of low order are ordinarily not separated from the principal maxima and those of high order are too faint to be seen.

(The foregoing conclusions apply to the subsidiary maxima formed by a grating which is perfect. When periodic error of ruling exists, subsidiary maxima are formed which are known as "ghosts." These will be discussed at the end of the chapter.)

It is of interest to compare the patterns which arise from each of two gratings one of which has n lines and the other $2n$ lines. We shall discuss two cases:

1. When the grating space is the same in the two gratings: In this case the width of the second grating must be twice as great as that of the first. The patterns then compare as the first and second patterns of Figure 113.

The principal maxima maintain their position, but both principal and subsidiary maxima are narrower in proportion as the grating is wider.

2. When the $2n$ lines are crowded into the original width $w=na$ of the first grating: In this case the patterns compare as the first and third patterns shown. The widths of both principal and subsidiary maxima are the same in the two patterns, but since in the second grating the grating space is half as great as in the first, the separation of the principal maxima is doubled.

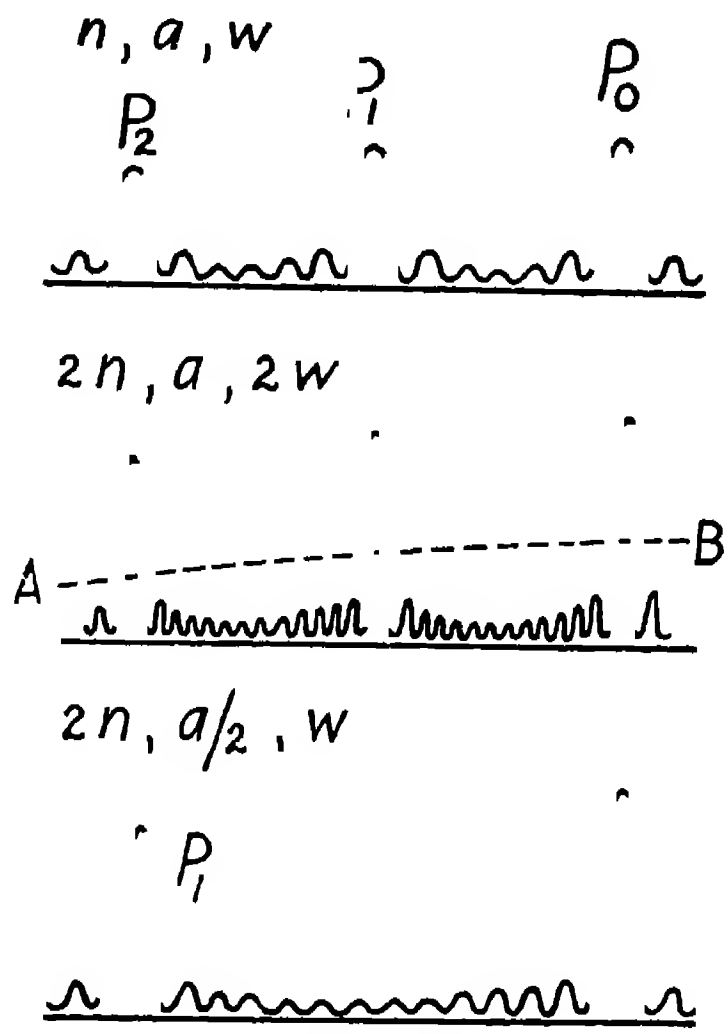


FIG. 113

We shall conclude this section by deriving a well-known equation for the resultant amplitude at any point of the diffraction pattern formed by a grating, in terms of the amplitude contributed by each line:

The amplitude contributed by each line varies with the direction of diffraction, as illustrated, e.g., in Figure 113 by the variation in height of the broken curve AB. Denote the amplitude for one line in a given direction by S_1 and denote by α the phase difference between disturbances from adjacent lines. Our problem consists in evaluating the resultant amplitude, which we shall denote by S, in terms of S_1, α, and n, the number of lines of the grating. The disturbance from each line will contribute one vector to a polygon (see Fig. 114). In this figure n is taken as a small number and the polygon is represented as making only a fraction of a convolution. The equation to be derived is, however, perfectly general—it applies for n either large or small and for a polygon making any number of convolutions. The angle between adjacent vectors is α. The polygon, being equilateral and equiangular, may be inscribed in a circle. Denote the radius of the inscribing circle by ρ. The angle subtended by one vector at the center of the circle is equal to α, and

$$\frac{S_1/2}{\rho}=\sin\frac{\alpha}{2}\,.$$

Now ψ, the angle indicated in the figure, equals $n\alpha$, and hence

$$\frac{S/2}{\rho}=\sin\frac{\psi}{2}=\sin\frac{n\alpha}{2}\,.$$

Eliminating ρ from these two equations:

$$S = S_1 \frac{\sin \frac{n\alpha}{2}}{\sin \frac{\alpha}{2}} . \tag{3}$$

This is the equation which was sought.*

In any given set-up the angle of incidence is constant and the amplitude S may be regarded as a function of the angle of diffraction. By introducing the general relations $\alpha = 2\pi\delta/\lambda$ and $\delta = a(\sin\theta_1 + \sin\theta_2)$, where δ is the path difference by way of adjacent lines of the grating, we may throw equation (3) into a form in which S is given as a function of the angle of diffraction, θ_2, explicitly. For example, at normal incidence $\sin\theta_1 = 0$ and thus $\alpha = 2\pi a \sin\theta_2/\lambda$. This value of α may now be substituted in equation (3). The resulting expression for S will however still contain S_1, the amplitude due to one line of the grating.

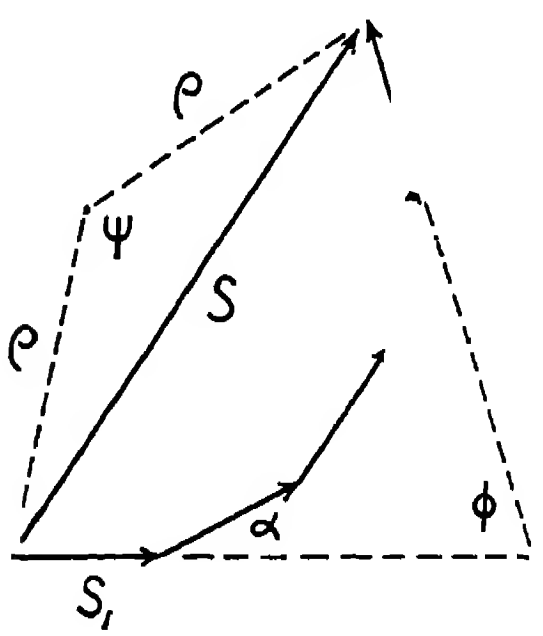


FIG. 114

The amplitude S_1 itself varies with the angle of diffraction; it, however, varies only slowly when the slits of the grating are narrow. The variation of S_1 can be taken into account when, and only when, we know the relative width of the transparent spaces and the opaque ones. The equation which then results will be derived at the end of section 15.

7. Periodic Structure as Prime Requisite.—We have thus far considered only the grating consisting of narrow slits in an opaque screen. However, any structure which is periodic in space—any structure which repeats itself at some small interval a—will serve as a diffraction grating. We can understand this if we consider first a wide rectangular aperture, without structure, and then consider this to be replaced by a periodic structure. Consider the simple aperture of width w to be divided into n strips each of width a so that $na = w$. We should expect a principal maximum of diffraction to be formed in a direction such that the path difference by way of corresponding points of these adjacent strips is λ (see upper portion of Fig. 115) were it not for the fact that in this direction the vibration curve for each strip closes in one convolution as indicated. In other words, when we have no structure we have no grating.

A principal maximum will be formed if we can by any method open up the vibration curves for all of the strips in the same manner. We shall then have a grating. In a grating consisting of an opaque screen with slits in it,

* The phase in which the resultant disturbance arrives is retarded with reference to that from the first line of the grating by $(n-1)\alpha/2$. This follows from the fact that the angle indicated in the figure by ϕ is equal to $\pi - (n-1)\alpha$.

the required opening-up of the vibration curves is accomplished by blocking off a portion of the wave front in each strip. We may, however, achieve the same result in a different manner. We may introduce changes in the phase which normally exists upon the wave front, at regularly spaced intervals across the wave front. Suppose, for example, that grooves are ruled at an interval a in a plate of glass (see lower portion of Fig. 115). Owing to modifications in the length of optical path now introduced, the vibration curve for each strip of width a will in general be an open one as indicated. We now have a principal maximum formed—we have a periodic structure and as a result we have a diffraction grating. The figure is drawn supposing that the ruled glass plate is used by transmission.

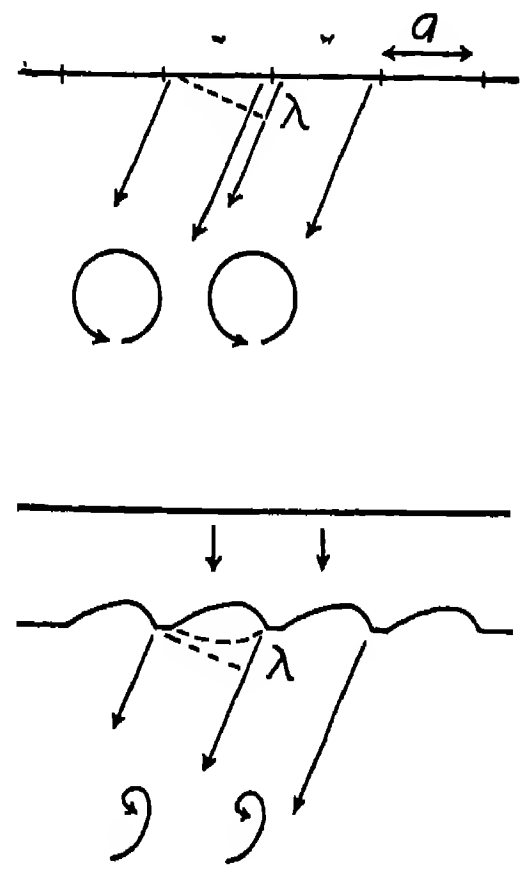


FIG. 115

8. The Reflection Grating.—The principle of the reflection grating is now almost self-evident. A groove ruled into a reflecting surface will likewise in general give rise to an open vibration curve. Referring to Figure 116, when the light is incident at an angle θ_1 with reference to the normal to the grating, the principal maximum of the mth order will be formed in that direction θ_2 for which the path difference by way of corresponding points of adjacent grooves equals $m\lambda$. The path difference is $ND+DE$. Since $ND=a \sin \theta_1$ and $DE=a \sin \theta_2$, we have $m\lambda=a \sin \theta_1+a \sin \theta_2$. Thus the same law holds for both the transmission grating and the reflection grating.

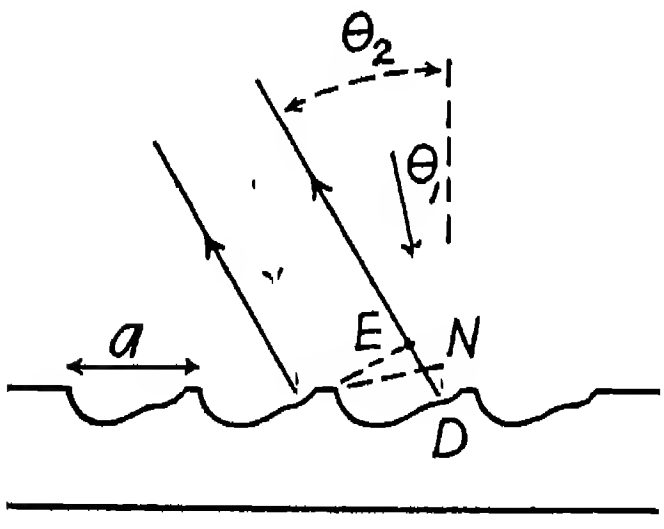


FIG. 116

9. Depth and Form of Grooves.—In ruling a grating it is aimed to rule with the proper depth to have the grooves practically contiguous, leaving little or no unruled or original surface between grooves. Such original surface when present reflects a maximum of light into the central image at the expense of intensity in the spectral images where it is needed. The form of the groove determines the direction into which the light will be predominantly thrown. If, for example, one face of the groove is flat, a maximum of energy will be thrown into the direction which makes an angle with this face which is equal to the angle made by the incident light with this face. The spectral order which lies in this general direction will be intense.

Great care is necessary in choosing a diamond for ruling. An edge is selected which appears under the microscope to be clean, that is, not rounded off. With this edge, trial rulings are made, and from these it can be ascertained whether a sufficient amount of light is diffracted into the desired direction. Natural edges of the diamond were used until recently. Within the last few years the grinding of diamonds for ruling has proved successful. The faces are usually ground to make a right angle with the hope that the groove faces will then make a right angle. But it does not follow that they will do so. The faces of the grooves may not even be flat. The ruling is only, let us say, one twenty-thousandth of an inch deep and the edge may not be clean to the required degree.

It is not possible to ascertain the actual form of the finished groove. Examination of the grating under a high-power microscope clearly reveals the individual grooves but shows little regarding their form. However, in coarse gratings, used for the infra-red, a fair approximation to flatness of groove face is undoubtedly attained (see "The Echelette," sec. 22). Numerous microphotographs of gratings have been published by L. C. Glaser.*

10. The Ruled Grating; Historical.—The diffraction grating was originally discovered in the year 1785 by the American astronomer David Rittenhouse,† but he did not follow up his work and his discovery appears never to have attracted general attention. The grating was rediscovered by Joseph Fraunhofer,‡ who announced his discovery in the year 1821. Fraunhofer's first gratings were made by winding fine wire over a rectangular frame which had a screw with a fine thread on each of two opposite edges. The threads served to space the wires equally. By this method Fraunhofer constructed gratings of which the number of "lines" per Paris inch§ ranged from 40 to 340. He despaired of making screws with more than 340 threads per inch, and, desiring to work with gratings of as small a grating space as possible, he next turned to the process of ruling.

Fraunhofer never described the machine which he used, but it was probably a "dividing machine"—a machine for dividing linear measuring scales—which he adapted for his purpose. Such machines are still used somewhat for ruling small gratings. Fraunhofer first ruled through gold foil laid on glass, up to nearly 900 lines per inch, and was successful up to about 1800 lines per inch in ruling in a thin film of grease on glass. The gratings were all small. Finally, using a diamond as a ruling point, he ruled directly into the surface of glass. One of his glass gratings had as high as 8000 lines per

* *Zeit. f. Tech. Physik*, **7**, 31, 1926.

† *Trans. Amer. Phil. Soc.*, **2**, 201, 1786; reprinted in *Jour. Franklin Inst.*, **214**, 99, 1932.

‡ *Gesammelte Schriften*, p. 67.

§ One Paris inch = 1.06577 inches = 2.70700 cm.

inch. This one was about ½ inch wide. It rendered him valuable service in verifying the grating law and in measuring the wave-lengths of the Fraunhofer lines in the solar spectrum.

The ruling of a satisfactory grating involves not merely the placing of a multitude of fine lines close together—in addition the grating space must be constant to within small permissible limits. Fraunhofer fully realized this, and while he succeeded in ruling separate lines even at the rate of 32,000 per inch, the grating space was then no longer constant with sufficient approximation. He estimated that the grating space would have to be constant to within 1 per cent, or, in a grating of this fineness, to within 1/3,200,000 part of an inch, and his machine was not capable of such performance. We shall learn when we come to consider "Error of Ruling" that the magnitude of the permitted errors depends markedly upon whether the errors are systematic or non-systematic in their character, but the precision which is required is always enormous.

Fraunhofer in his spectroscopic investigations used his gratings mostly by transmission. He, however, also discovered the principle of the reflection grating and showed that the grating law holds in either case.

About twenty-five years elapsed after Fraunhofer's discovery of the grating before anyone undertook to furnish gratings for scientific use. Then F. A. Nobert, an instrument maker of Pomerania, a province of Prussia, began to offer them. The majority of Nobert's gratings were ruled by a diamond directly on glass, but some were ruled through silver on glass. They were ruled for the most part at either 3000 or 6000 lines per inch. These remained the best available for another period of twenty-five years.

In about 1870 L. M. Rutherfurd, a New York lawyer, became interested in the ruling of gratings by way of an avocation. After a few years he succeeded in producing gratings far superior to those of Nobert. He, moreover, introduced the speculum reflection grating. Speculum is an alloy of 68 per cent copper and 32 per cent tin. It is hard, yet softer than glass. It can be ground and polished much like glass, and combines high reflecting power with resistance to tarnish. Most of Rutherfurd's gratings had about 17,300 lines per inch. They were from 1 to 2 inches square. His gratings were the first to surpass the most powerful prisms in performance.

We owe the diffraction grating of today to the genius of Henry A. Rowland, who was professor of physics at the Johns Hopkins University from 1876 until his death in 1901. Rowland made two great advances. In the first place, he constructed vastly improved ruling machines, and was thus able to rule gratings of much greater size and perfection than his predecessors; and, in the second place, he invented the *concave* grating, which is of tremendous scientific value. Rowland, developing both theory and prac-

tice, showed that a grating ruled upon a concave surface may be used to form a sharply focused spectrum without the use of lenses. The concave grating will be dealt with in later sections and its value will then be pointed out.*

By far the majority of all gratings which have been ruled up to the present have been ruled upon three machines which were constructed by Rowland, and operated at first by Rowland assisted by Schneider, then after Rowland's death by J. A. Anderson, and recently by R. W. Wood. Ruling machines have also been constructed and operated at the University of Chicago, at the Mount Wilson Observatory, and at the Bureau of Standards. A few excellent gratings have been ruled upon these. There is a machine for ruling coarse gratings at the University of Michigan upon which very effective gratings for work in the infra-red have been ruled by E. F. Barker. In England the Blytheswood machine at the National Physical Laboratory at Teddington is suitable for ruling gratings of medium size. There are numerous machines in existence, including dividing machines, which are suitable for ruling small gratings. Additional machines for ruling large gratings are in process of construction.

11. Rowland's Ruling Machines.—Figure 117 may be considered to represent one of Rowland's machines schematically. It is not accurate in detail.† The plate which is in process of being ruled, *1*, rests upon the "grating carriage," *2*. The diamond, the ruling point, is fastened to the lower end of a small rod, *3*. During the "ruling stroke" of the machine the diamond rests on the plate, *1*, and is pulled over the surface by motion of the "ruling carriage." This carriage has the form of a bridge or span extending over the grating from *4* to *5*. The ruling carriage is sometimes called the "diamond carriage." The rod *3* is mounted in a mechanism which is suspended from the span. This mechanism involves a swivel joint having a horizontal axis. Each end of the ruling carriage rests upon a pair of parallel ways which extend from *6* to *7* except for a lacking middle section. During the ruling

* Rowland's writings were collected after his death in a volume entitled *Rowland's Physical Papers* (Baltimore: Johns Hopkins Press, 1902). The first paper which Rowland wrote on the diffraction grating was a brief announcement which told of the improvement in ruling which he had been able to achieve and also told of his discovery of the focusing properties of the concave grating. This announcement was made in 1882. It appeared in several journals, e.g., *Phil. Mag.* (4) **13**, 369, 1882.

† The mechanism of the actual machine is intricate. The figure attempts to present the principal features of the machine clearly in a single drawing. To further this end omissions and deliberate misrepresentations of detail have been necessary.

Rowland never published a description of his machines. A description with numerous scale drawings was, however, prepared after Rowland's death under the direction of J. S. Ames and incorporated at the end of *Rowland's Physical Papers*. A description by J. A. Anderson is given in Glazebrook, *Dictionary of Applied Physics*, **4**, 30.

stroke of the machine the ruling carriage is pulled along these ways from right to left by the connecting rod *17* which is connected at its left end to a crank, *16*, mounted on the driving shaft, *13*. This shaft turns in the direction indicated by the arrow, being driven by a motor through a light belt and pulley omitted from the drawing. The driving shaft also has upon it

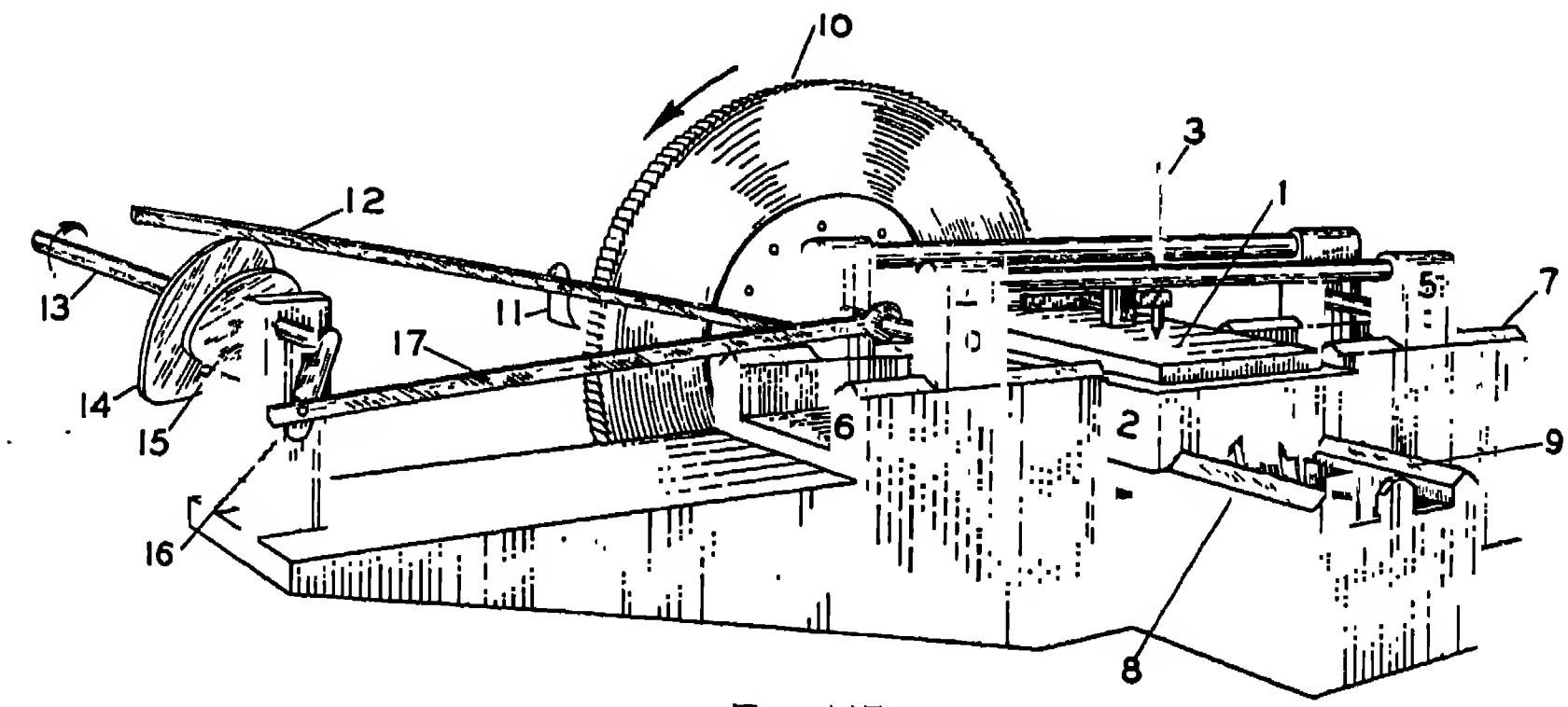


FIG. 117

the cams *14* and *15*. When the ruling of one line or groove has just been completed, the ruling carriage is at the left end of its range of travel. The cam, *15*, then raises a lever which rests upon it (omitted from the drawing). This lever in turn operates a mechanism which raises the rod *3* and thus raises the diamond from the surface of the grating and holds the diamond in raised position during the "return stroke" of the machine in which the ruling carriage is pushed by the connecting rod, *17*, to the right-hand end of its range of travel. Furthermore, during the return stroke the grating carriage, *2*, and with it the grating are given a minute displacement along the parallel ways *8* and *9*, which are perpendicular to the ways along which the ruling carriage travels. Beneath the grating carriage and connected to it by a special mechanism is a rather long nut which does not show in the figure. This nut encircles the main screw of the ruling machine, the "feed screw." The fore part of this screw is shown, lying between the ways *8* and *9* and running parallel to them. At the far end of the feed screw and rigidly connected to it is a large-toothed wheel, the "spacing wheel," *10*. The lever *12* is raised during the return stroke by the cam *14*, and the "pawl," *11*, is caused to tip to the right at its lower end by a mechanism not shown, and engage in a tooth of the spacing wheel. The lever *12* is then allowed to fall again, still during the return stroke, and thus the spacing wheel and the feed screw are given a small rotation. The rotation of the feed screw causes the nut to advance along the screw a minute distance and carry with it the grating carriage and the grating. The mechanism which operates the pawl,

11, is a lever, adjacent to *12*, also resting on the cam *14* which is really a double cam. The operation of turning the spacing wheel, and thus displacing the grating laterally, takes place during the return stroke of the machine while the diamond is in raised position. As this operation draws to completion the ruling carriage reaches the right-hand end of its range of travel and the diamond is allowed to descend gently upon the grating surface. All is now in readiness for the ruling of the next line. The lines are thus ruled one by one until the grating is complete.

The ruling of a large grating requires from three to five days of continuous operation. Even a slight change in the speed of the machine is disastrous. The machine must be mounted where there is freedom from earth tremors, and temperature must be maintained constant to within a hundredth of a degree. Before starting, the room is closed; the machine is started from the outside, and no one enters until the grating has been completed.

Rowland's machines have a capacity for ruling gratings of over 6 inches in width with a length of ruled line of over 4 inches. However, few gratings have been ruled to either this full width or with the full length of line.

The feed screws are 17 inches long and 1 inch in diameter; they originally all had twenty threads per inch, very nearly. The feed screw is an important element of the machine and it was largely through devising a method of making screws practically free from error that Rowland was able to perfect the diffraction grating.*

The spacing head of Rowland's first machine has $2 \times 360 = 720$ teeth, or one tooth for every half-degree of arc. Thus when the machine is set to use every tooth on the spacing head (and not every other tooth, or every third tooth), the number of lines ruled per inch would be $20 \times 720 = 14{,}400$, if the pitch of the screw were exactly one-twentieth of an inch. Actually the first machine rules about 14,438 lines per inch. One of the other machines has a spacing head with 1000 teeth. This originally ruled 20,000 lines per inch almost exactly, when every tooth was used, or 10,000 lines per inch when every second tooth was used, or 6667 lines per inch when every third tooth was used, etc. In about the year 1913, Anderson substituted a new screw in this machine which made the fundamental spacing 30,000 lines per inch. The third machine rules gratings of 15,000 lines per inch.

12. Replicas.—It is possible to make copies or "replicas" from gratings ruled upon glass or speculum. Excellent replicas can now be obtained and they are very much less expensive than original, ruled gratings. Replicas

* The innovations introduced by Rowland consisted principally in making the screw considerably longer than it is finally to be, and in grinding with emery by running a quite long nut back and forth over the screw, and then polishing with rouge in the same manner. The process requires several weeks; details of technique are elaborate (see *Rowland's Physical Papers*, p. 505, or J. A. Anderson in Glazebrook, *op. cit.*, **4**, 30).

are commonly used for instructional purposes and, when of good quality, they are amply good also for certain phases of investigational work.

Among the early experimenters upon the making of replicas are to be mentioned particularly John Strutt, who later became Lord Rayleigh,* and G. Quincke.† They both made their first reports in the year 1872. Rayleigh continued his experiments at intervals for a number of years. His method consisted in making photographic contact prints, mainly from gratings on glass. He also tried making casts by pouring filtered gelatin over a grating, and when the gelatin had dried, stripping it off. He attributes this method to Brewster. He, however, rejected it in favor of the photographic method.

Quincke deposited a thin film of silver chemically upon a glass grating, then thickened the coating by electroplating and finally stripped it off and mounted it upon a flat plate of glass, thus producing a reflection grating with a silver surface. These replicas, however, lacked permanence owing to tarnishing of the silver. Quincke also experimented with the making of casts, using collodion instead of gelatin.

T. Thorp‡ appears to have been the first who achieved real success by the method of casts. He used celluloid. This method was subsequently developed by R. J. Wallace§ and by F. E. Ives.||

The excellent replicas of the present day are casts made by Wallace. His method is in outline as follows: Guncotton (pyroxilin) is dissolved in amyl acetate, and a small quantity of the resulting collodion is poured over the grating and allowed to dry slowly. Thereupon the grating with the cast still adhering is placed in a tray of distilled water. When the cast shows signs of loosening from the grating it is stripped off and mounted, face up, on a plate of glass which has been previously covered with a thin layer of gelatin. The cast is again allowed to dry, whereupon the replica is ready for use. There are many details of technique to be observed which we cannot enter into here. The casts shrink somewhat, and it is difficult, and a matter of prime importance, to secure even shrinkage.

Reflection replicas, either plane or concave, are made by placing the cast, face down, on a previously silvered flat or concave glass base. The collodion film protects the silver from tarnish.

* *Scientific Papers*, **1**, 157, 160, 199, 504, and **4**, 226.

† *Pogg. Ann.*, **146**, 1–65, 1872; in particular pp. 33 and 44.

‡ British Patent No. 11466 (1899). This patent is abstracted in *Photography*, an English journal later continued as *Photography and Focus*, **12**, 514, 1900. On p. 121 of the same volume it is claimed that a description of Thorp's method had been previously given in this journal. A cursory search failed to reveal this. No other description appears to have been published.

§ *Astrophys. Jour.*, **22**, 123, 1905.

|| *Jour. Franklin Inst.*, **159**, 457, 1905.

An alternative method of making reflection replicas which appears, however, never to have been extensively used, is to place the cast, face up, on an unsilvered base, and then cover it with a thin film of a suitable metal, either by cathode sputtering or by evaporation.

13. Crossed Gratings.—When a small source of light is viewed through two transmission replicas held in contact, with the lines of one replica let us say vertical and the lines of the other horizontal, one sees the spectra which would be formed by each grating acting alone, and in addition one sees diagonal spectra which arise from the combined action of the two gratings.

The formation of spectra by crossed gratings takes place in the same manner as does the formation of spectra by a so-called "plane net." The diffraction by a "plane net" is of importance in connection with the diffraction of X-rays by crystals, and the whole question will be treated in detail in the chapter on X-rays, chapter 8, sections 14 and 15. We shall for the present content ourselves with accounting in a general qualitative manner for the spectra which are formed. The reader may do well to refer in this connection to chapter 8, Figure 212, which represents diagrammatically the array of spectra which occur in one quadrant. The plane of the figure is that of the photographic plate, or, when we are observing visually, of the retina of the eye.

Suppose the source contains three wave-lengths, λ', λ'', and λ''', such that $\lambda' < \lambda'' < \lambda'''$. The first grating, with lines vertical, will diffract these three wave-lengths, in a given order, m, at three distinct horizontal angles, $\alpha' < \alpha'' < \alpha'''$. A part of these beams will pass without further deviation through the second grating, thus accounting for the horizontal spectra formed by the first grating. But a part of each of the three beams will be diffracted vertically in a given order by the second grating at three angles $\beta' < \beta'' < \beta'''$. The resulting spectrum will evidently lie on a diagonal. If the two gratings have the same grating space and the light is incident normally, each angle β will be equal to the corresponding angle α when the order of vertical diffraction is the same as that of horizontal diffraction. The spectra, having the same order of diffraction both times, will lie on the 45° diagonal. By considering in succession each of the horizontal orders, *1*, *2*, etc., and for each of these considering the vertical orders *1*, *2*, etc., all of the diagonal spectra may be accounted for. It is perhaps needless to add that the vertical spectra are formed by the second grating out of the light which has passed without deviation through the first grating.

14. Wire Gratings; Gratings Ruled through a Silver Film.—The wire grating, which is historically the oldest form of grating, is still used occasionally for investigation in the far infra-red (40–300 μ). In addition, this form of grating holds a general interest because, being composed of opaque

and transparent spaces of known relative widths, it permits the calculation of the relative intensities of the various spectral orders. The important feature of the transparent spaces is really not so much that they are transparent but that they are what we shall call "clear." That is, assuming perpendicular incidence, the phase of the disturbance is constant across each clear space in the plane of the grating, and is the same in all of the spaces. This same feature is presented by gratings formed when a deposit of silver on a plate of glass is ruled away in strips. We may place these two types of gratings in one category, namely, of "gratings composed of opaque bars and clear strips."

The most important special case is that in which the width of the bars equals the width of the clear strips. In making wire gratings this equality of width is achieved by following the method of Du Bois and Rubens.* The method consists in winding two strands of fine wire side by side upon a metal frame and then unwinding one of the strands. After soldering the remaining strand to the frame, the wires on the back of the frame are cut away, leaving the desired grating.†

15. Intensity Ratios; Absent Spectra.—When a grating has clear strips equal in width to the opaque ones, the spectra of even order are entirely absent and the odd orders decrease rapidly in intensity with increasing order. Referring to Figure 118, suppose that the amplitude contributed at the central image by one clear strip is represented by the vector A_1 and that the broken curve, B, represents the pattern due to one clear strip acting alone (amplitude of illumination). The first minimum in the curve B occurs when the vibration curve for one

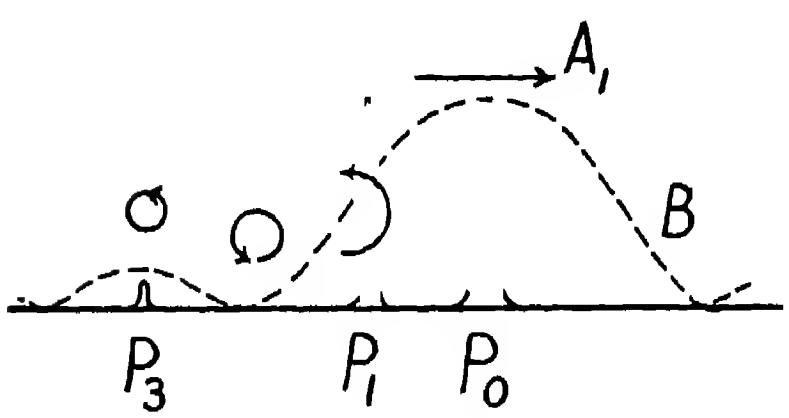


FIG. 118

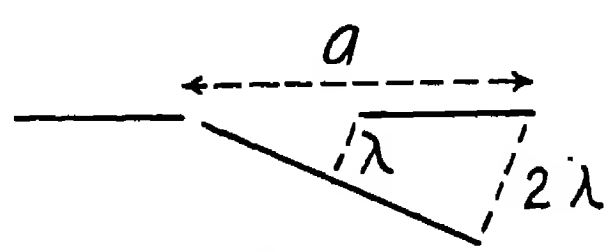


FIG. 119

clear strip closes in one convolution as indicated. At this point the path difference by way of the right- and left-hand edges of the strip is λ (see Fig. 119) and the path difference by way of corresponding points in adjacent clear strips is 2λ. The first minimum in the broken curve is therefore

* *Wied. Ann.*, **49**, 593, 1893.

† The frame should be made in such a way that it can be slightly expanded by means of screws after the grating is complete, in order to draw the wires taut.

the point at which the grating should form its second-order principal maximum, but this maximum is not formed because the contribution from each clear strip is equal to zero at the point in question. Similar reasoning shows that the other even orders are also absent.

Gratings other than those composed of opaque and clear strips occasionally exhibit the phenomenon of suppressing certain spectral orders—or perhaps only a part of the spectrum of a given order. We usually have little or no information about the groove form in such cases and can only say that under the given conditions the disturbances from the individual grooves must be zero, otherwise the order would have to be present.

Returning to our grating composed of equal opaque and clear strips, we shall compute the relative intensities of the central image, P_0, Figure 118, the first-order image, P_1, and the third-order image, P_3. If the grating has n "lines," the intensity of P_0 is $(nA_1)^2$. At P_1 the length of arc A_1 forms a semicircle which has a closing vector of length $2A_1/\pi$. The intensity at P_1 is hence $(2A_1n/\pi)^2$, or less than that at P_0 by a factor $4/\pi^2=4/10$. At P_3 the length of arc A_1 forms a circle of one and one-half convolutions. It has therefore a closing vector of length $2A_1/3\pi$ and the intensity is $(2A_1n/3\pi)^2$, or less than that at P_0 by a factor of $4/9\pi^2=4/90$ approximately.

It should be mentioned that with wire gratings reflection of light at the surface of the wires may cause the actual intensities to differ somewhat from the calculated values.

When a grating, of grating space a, has clear strips each of width $a/3$ and opaque strips of width $2a/3$, then it may be shown by a process similar to that followed above that the third, sixth, and ninth orders will be absent. Moreover, it may be shown that these same orders will be absent when a grating has clear strips of width $2a/3$ and opaque strips of width $a/3$. The two gratings in question constitute complementary diffraction screens and not only are the same spectra absent for each of the two gratings but the intensities of the spectra which are present are the same for the two gratings, order for order, and not only relatively, but absolutely. Only the central images and the subsidiary maxima in their immediate vicinity differ in intensity. This can readily be proved by drawing the appropriate vibration curves. The fact follows also from *Babinet's principle*. This principle, which is treated in chapter 7, section 12, enables us to draw certain broad conclusions regarding the patterns formed by complementary diffraction screens.

In conclusion we shall treat the case of a grating composed of clear and opaque strips of any relative width:

Denote the grating space by a, the width of the clear strips by b, and the width of the opaque strips by c.

CASE A. Intensity of the *spectral images* (spectral order, m, integral).—Disturbances traveling by way of corresponding points of adjacent clear strips arrive at the

spectral image of order m with a phase difference of $\alpha = 2\pi m$. Consequently, disturbances traveling respectively by way of the right- and left-hand edges of any one clear strip arrive with a phase difference of $\beta = 2\pi mb/(b+c)$. For each clear strip we have a vibration curve of length of arc A_1, and this curve has a closing vector of a length S_1 which can be computed by equation (2) of chapter 5, which, written in our present notation, becomes:

$$S_1 = A_1 \frac{\sin \frac{\beta}{2}}{\frac{\beta}{2}} = A_1 \frac{\sin \frac{\pi mb}{b+c}}{\frac{\pi mb}{b+c}} . \tag{4}$$

Here S_1 is the amplitude of the resultant disturbance from one clear strip. The intensity of the spectral image in question is proportional to $(nS_1)^2$, where n is the number of lines of the grating. For a given grating n is constant and we may represent the intensity simply by S_1^2. Correspondingly, the intensity of the central image becomes A_1^2. We may now write $A_1^2 = [A_0 b/(b+c)]^2$ where A_0^2 is the intensity which would prevail at the central image if the entire ruled area of the grating were transparent. Squaring equation (4) and then substituting the given value of A_1, the intensity of the spectrum of order m becomes:

$$S_1^2 = \frac{A_0^2}{\pi^2 m^2} \sin^2 \frac{\pi mb}{b+c} . \tag{4'}$$

Since the sine function can never be greater than unity, the greatest proportion of the original energy which can go into the spectral image of order m is $1/\pi^2 m^2$. Moreover, the first-order image is always the most intense in the type of grating under consideration, and this image cannot have more than $1/\pi^2$ or about the one-tenth part of the original light. In order to have this intensity we must have the sine equal to unity—or, in other words, $b = c$; that is, the width of the transparent spaces and the width of the opaque ones must be equal.

Recapitulating, for a grating having opaque and transparent strips of equal width (see preceding paragraph and also beginning of section):

If we place the intensity of the original beam	$=1$
then, the intensity of the central image	$=\frac{1}{4}$
the intensity of the first order	$=\frac{1}{\pi^2} = \frac{1}{10}$ approx.
the intensity of the second order	$=0$
the intensity of the third order	$=\frac{1}{9\pi^2} = \frac{1}{90}$ approx.

All of the foregoing holds only for gratings composed of opaque and clear strips. Gratings of other types may throw a distinctly larger proportion of the original energy into one spectrum, and not always the largest proportion into the first-order spectrum.

CASE B. Amplitude at *any field point* (spectral order, m, in general fractional).—The n clear strips of the grating contribute n vectors each of length S_1 to a vibration

polygon. The closing vector of this polygon has a length S given by equation (3) which we shall repeat here:

$$S = S_1 \frac{\sin \frac{na}{2}}{\sin \frac{a}{2}} . \tag{3}$$

In this equation a denotes the angle between successive vectors of the polygon—or, in other words, the phase difference between disturbances arriving by way of corresponding points of adjacent clear strips. The phase difference a equals $2\pi m$, where m, the order, is integral when we are at one of the principal maxima of diffraction, but at a general point of the pattern, m will have a fractional value, namely, $a/2\pi$. We obtain our desired equation by substituting in equation (3) the value of S_1 from equation (4), at the same time writing $\pi m = a/2$. That is:

$$S = A_1 \frac{\sin \frac{ab}{2(b+c)}}{\frac{ab}{2(b+c)}} \frac{\sin \frac{na}{2}}{\sin \frac{a}{2}} . \tag{4''}$$

The value of A_1 may be expressed in terms of A_0 if desired, that is, we may substitute $A_1 = A_0 b/(b+c)$.

In deriving equation (3) it was pointed out that this equation could be transformed so that S is expressed as a function of the angle of diffraction θ_2, instead of as a function of a. Equation (4″) admits of similar transformation.*

16. Polarization.—The light diffracted by a grating is always polarized to a greater or less degree.

We may consider the incident, unpolarized light to be resolved into two components: one component having the electric vector parallel to the grooves and the other component having the electric vector perpendicular to the grooves. The presence of polarization in the emergent beams means that the parallel component of the incident beam is distributed by the grating among the central image and the spectral images in one proportion of intensity and the perpendicular component is distributed in different proportions.

The polarizing effects of gratings were discovered by Fraunhofer, who found, when he was using one of his glass gratings by reflection, that at a certain angle of incidence the diffracted light of a given wave-length and

* Formulae giving the amplitude or intensity, for gratings of several types, including gratings composed of opaque and clear strips for which the formulae are here given, were developed during the early seventies principally by Lord Rayleigh and by G. Quincke (see Rayleigh, *op. cit.*, **1**, 199, or the article "Wave Theory" in *Encyc. Brit.* [9th ed.], **24**, 421, and Quincke's long series of articles, "Optische Experimental-Untersuchungen," in *Pogg. Ann.*, e.g., article No. 15 of the series in **146**, 1, 1872).

spectral order was completely polarized. As the angle of incidence varied, the wave-length at which complete polarization occurred was found to vary.

If in Fraunhofer's experiment the *incident* light had been polarized at right angles to the azimuth in which he found the diffracted beam to be polarized, then the diffracted spectrum would evidently have had a dark band at the wave-length in question. A part of one spectral order would have been absent. Such local suppression has since been found when properly polarized incident light is used.

A detailed discussion of the polarization effects produced by gratings would carry us too far afield. Attention having been called to the existence of such effects, the reader is referred for further information to the sources mentioned in the footnote.*

One of Hertz's numerous experiments with electromagnetic waves consisted in showing that a grating formed of parallel wires is almost opaque to electromagnetic waves when the electric vector is parallel to the wires, but is almost transparent to them when the electric vector is perpendicular to the wires. This experiment is related to the polarization effects produced by optical gratings. However, in Hertz's experiment the wave-length is many times as great as the grating space, whereas in the optical case the wave-length is usually smaller than the grating space.

The effects found in the central image when fine wire gratings are used optically are the reverse of those found in Hertz's experiment.

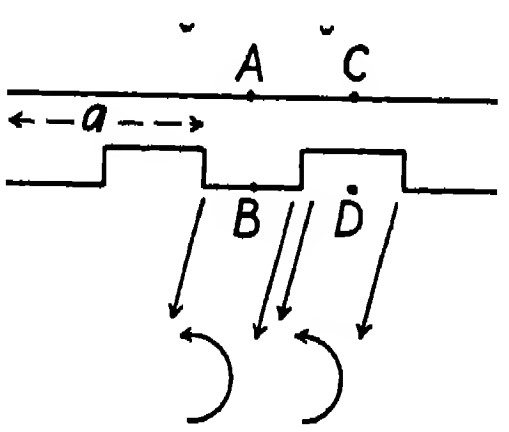


FIG. 120

17. The Laminary Grating.—A grating of the type represented in cross-section in Figure 120 is known as a laminary grating.† Let us suppose that the grating is a transmission grating of glass and that monochromatic light of wave-length λ is incident. If the optical path AB, which lies entirely in glass, exceeds the optical path CD, which lies partly in air, by an amount $\lambda/2$, then the central image will be suppressed because disturbances from the raised portion at B will arrive out of phase with the disturbances from the depressed portion at D. With a given grating this holds only for a given wave-length. The light which would normally go into the central image is thrown into the diffracted images. In the first order, the vibration curve for

* Du Bois and Rubens, *Annalen*, **49**, 593, 1893, and *Berichte*, **2**, 77, 1904. R. W. Wood, *Phil. Mag.* (6), **4**, 396, 1902, and *Physical Optics* (Macmillan, 1914), p. 632; Geiger and Scheel, *Handb. d. Phys.*, **20**, 306; Lord Rayleigh, *op. cit.*, **5**, 388, 405, and 410. The aforementioned observations of Fraunhofer are described in his *Gesammelte Schriften*, ed. E. Lommel (1888), p. 134.

† The theory of the laminary grating was developed by Quincke, *Pogg. Ann.*, **132**, 364, 1867.

each raised portion is a semicircle, as indicated in the figure, and the vibration curve for each depressed portion is an identical semicircle. The intensity of the first order is hence four times as great as it would be if, for example, all of the depressed strips were opaque. Corresponding conclusions apply for the other odd orders. The images of even order are absent.

When a laminary grating is correct for sodium light, then a sodium flame cannot be seen through it, although every element of the grating is in itself clear. That is, the flame cannot be seen in its true position. Images of the flame, corresponding to the various odd spectral orders, are seen in deviated positions.

Reflecting laminary gratings have also been constructed. In these the depressions are one-quarter of a wave-length deep. Reflecting laminaries have recently been used in some investigations in the far infra-red.* The transmission laminary has found little or no application in spectroscopic work.

THE PLANE GRATING IN PRACTICE
(Secs. 18–22)

18. Direct Vision.—The simplest method of using a plane grating is to hold it before the eye and, supposing it is a transmission grating, to look through it at an illuminated slit. Referring to Figure 121, let S represent the source of which the spectrum is to be examined. Light from this passes through a slit, which we shall suppose to be vertical, and is incident upon the grating G, which is held with the lines vertical. A portion of, for example, the first-order diffracted beam of wave-length λ enters the pupil of the eye and is focused upon the retina, let us say at the point R. This beam appears to come from a point A, which is a virtual diffracted image of the slit. There will be other images in other wave-lengths and in other orders. Thus the observer sees spectra projected into space. These will occur symmetrically on both sides of the slit. The slit, seen in its true position, is itself the "central image" in this arrangement.

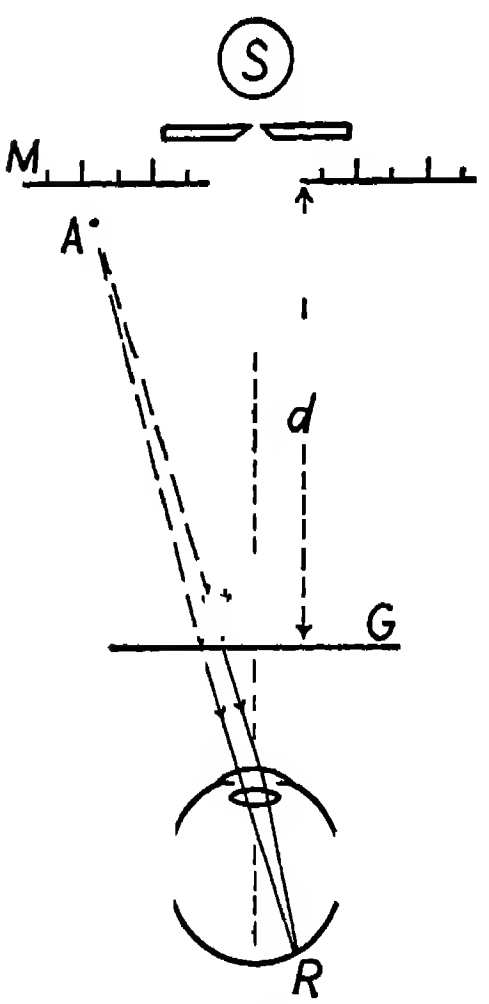


FIG. 121

The direction in which a given diffracted image lies is determined by the grating law. When it is desired to determine the wave-lengths of given spectrum lines by direct vision through a grating, it is most convenient to keep the angle

* Badger and Cartwright, *Phys. Rev.*, **33**, 692, 1929; C. H. Cartwright, *JOSA*, **21**, 784, 1931.

of incidence constant throughout. The grating should be placed in a support and the eye should be moved to one side or the other as may be necessary in order to view the various spectral images always through the same portion of the grating, let us say through the center. An index should be placed in contact with the grating to facilitate doing this. Then, when the light is incident perpendicularly, the grating law becomes $m\lambda = a \sin \theta_2$. The angle θ_2 can be determined by placing a meter stick, M, at some convenient distance, d, from the grating, just above or just below the line of sight, and then observing the position of a given image, A, as read on the meter stick. Denoting the distance from A to its counterpart on the right-hand side of the central image, again as read on the meter stick, by $2e$, we have $e/d = \tan \theta_2$, and, knowing the tangent, the sine can be found. The grating constant a must be known in advance or may be determined from a known spectrum line. Wave-length determinations made by direct vision through a grating lay no claim to high accuracy.

The distance of the image A from the point at which the beam penetrates the grating may be determined according to a formula which will be derived in discussing the concave grating (sec. 26, eq. [18]); the plane grating is a concave grating of infinite radius of curvature.

When a divergent beam of light falls upon a grating, as in the case under consideration, the virtual diffracted image A is well defined only while the effective aperture of the beam is not all too wide. When it is desired to use a wide beam and secure the best attainable definition, it is necessary to use a collimator to make the light parallel before it falls upon the grating.

19. The Common Spectrometer.—Figure 122 represents a transmission grating mounted in a spectrometer designed for visual observation. Light from the source S enters the slit of the instrument and passing down the collimator C leaves this as a parallel beam. The light is incident at an angle θ_1 upon the grating G. Parallel rays diffracted by the grating at an angle θ_2 enter the telescope and are brought to a focus on the crosshair within the tube at a point opposite P. Observations are made through the eyepiece E. The angles of incidence and diffraction are determined by reading the settings of the collimator and telescope with reference to the divided circle D above which the grating is mounted. The wave-length of a given spectrum line is calculated according to the grating law, $m\lambda = a \sin \theta_1 + a \sin \theta_2$.

The circular scale in the lower part of the figure represents a scale of wave-lengths. Each unit of the scale is to be taken as representing 1000 Ångstrom units of wave-length, the grating having supposedly a grating space of the proper magnitude for this to hold true. The position of the central image, zero wave-length, lies on the continuation of the axis of the collimator. The telescope is represented as being set at the position of about

6000 Å. It is to be noted that the wave-length scale is nearly uniform over the interval from 0 to 5. In this interval the angle of diffraction θ_2 is small. As the angle of diffraction increases the scale becomes less uniform, and beyond the point 7 the scale is markedly non-uniform. The divisions be-

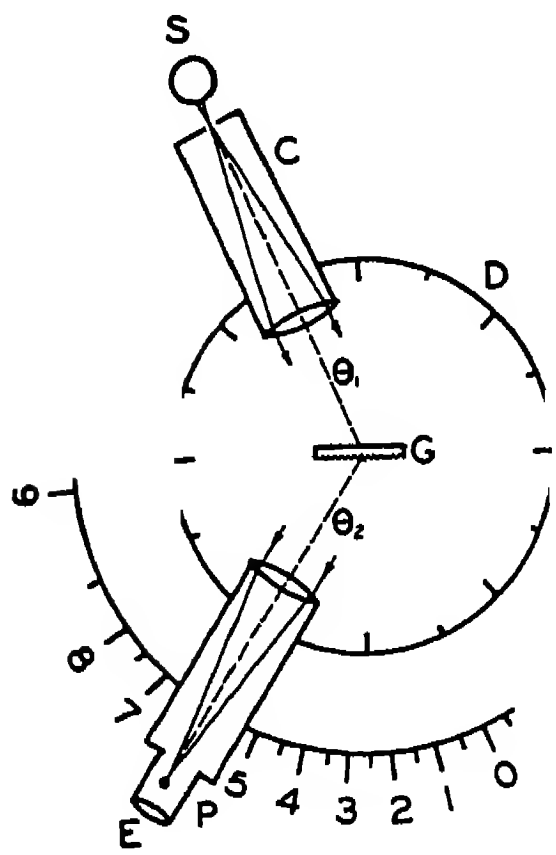


FIG. 122

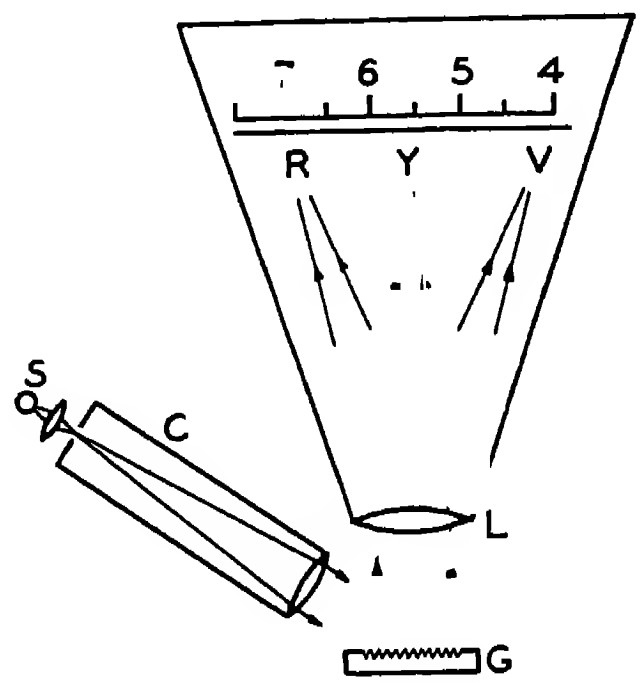


FIG. 123

FIG. 124

coming farther and farther separated, the *dispersion* increases enormously. It is readily seen why the scale has these characteristics. Differentiating the grating equation, we have:

$$m\,d\lambda = a\,(\cos\theta_1 d\theta_1 + \cos\theta_2 d\theta_2)\,.$$

When the angle of incidence θ_1 is constant, as here supposed, then $d\theta_1 = 0$. Consequently:

$$\frac{d\theta_2}{d\lambda} = \frac{m}{a\cos\theta_2}\,. \tag{5}$$

Now $d\theta_2/d\lambda$ is the *angular* dispersion. For directions of diffraction which lie near the normal to the grating, θ_2 is a small angle, and the cosine function varies slowly in the vicinity of zero angle. It is for this reason that the wave-length scale is nearly uniform from 0 to 5. If, by way of approximation, we take $\cos\theta_2 = 1$ over this interval, we have for the angular dispersion $m/2$, a constant. A spectrum in which the dispersion is constant over a given interval is said to be "normal" within this interval. It so happens that a grating spectrum is normal in the vicinity of the normal to the grating. This is a coincidence from the point of view of terminology but it is a coincidence worth bearing in mind. Equation (5) tells us further that the dispersion is proportional to the order of the spectrum, m, and inversely proportional to the grating space, a.

As the angle of diffraction θ_2 approaches 90°, cos θ_2 approaches zero, and the dispersion approaches infinity. It is for this reason that the divisions of our scale are far separated in the region of large values of θ_2. The dispersion of a grating becomes infinite in the grazing direction.

20. The Spectrograph.—Figure 123 represents a spectro*graph*, an instrument for *photographing* the spectrum. Light from the source, S, enters the slit of the spectrograph. A condensing lens is generally used for forming an image of the source on the slit, as here illustrated. The light passes down the collimator and is incident upon the grating G, which is represented as being a reflection grating. Parallel rays diffracted by the grating are focused by the camera lens L upon the photographic plate. A given spectrum line in the violet portion of the spectrum is focused, let us say, at V, a given line in the yellow at Y, and a given line in the red at R. When the camera faces the grating squarely, as here indicated, the linear dispersion is practically normal over the length of the photographic plate. To indicate this fact a uniform scale of wave-lengths has been inserted in the figure. If dx is an element of length taken along the photographic plate and f is the focal length of the camera lens, then $d\theta_2 = dx/f$ and the *linear* dispersion is:

$$\frac{dx}{d\lambda} = \frac{mf}{a \cos \theta_2}. \tag{5'}$$

In specifying the dispersion of a spectrograph, instead of giving the dispersion itself, the reciprocal of the dispersion, $d\lambda/dx$, is commonly stated, for example, 100 Ångstrom units of wave-length per centimeter along the spectrum.

A *spectrometer* is provided with an accurate circular scale, but a *spectrograph* is not usually provided with any accurate scale. Instead, a known spectrum, or comparison spectrum, is photographed in juxtaposition with the unknown spectrum in which the wave-lengths of the lines are to be determined.* When the dispersion is normal, if λ_1 and λ_2 are the wave-lengths of two lines in the known spectrum and x_1 and x_2 are the positions of these lines in the spectrogram, as measured upon a comparator, from an arbitrary zero, then $(\lambda_2 - \lambda_1)/(x_2 - x_1) = k$, a constant, and the wave-length λ of any spectrum line in the unknown spectrum is given by the relation $(\lambda - \lambda_1)/(x - x_1) = k$, where x is the position of the spectrum line in question. Whence:

$$\lambda = kx + (\lambda_1 - kx_1). \tag{6}$$

* A small 45° reflecting "comparison prism" is mounted just outside of the slit and the comparison spectrum is thrown on from the side, or a diaphragm with complementary apertures in it is arranged to slide over the face of the slit and the unknown spectrum and the comparison spectrum are photographed in succession.

The quantity in parentheses is constant and the calculation of the unknown wave-length is therefore a simple matter.*

The spectrograph of Figure 123 may be used with a transmission grating, in place of a reflection grating, by altering the relative positions of the collimator and the camera. Similarly, the spectrometer of Figure 122 may be used with a reflection grating.

Figure 124 represents a Littrow type of spectrograph. The distinguishing feature of this type is the fact that a single lens serves both for collimating and for focusing. Light from the source, S, is focused upon the slit of the spectrograph whence it passes to the lens, L, and thence to the grating. The slit lies just a little below the axis of the lens, or let us say just a little below the plane of our diagram; the incident beam consequently strikes the grating in a slightly upward direction. The diffracted beams are hence also directed slightly upward. The beam which is diffracted at the same angle, measured horizontally, at which the original beam is incident, is brought to a focus just above the slit, that is, above the plane of the figure. The photographic plate is situated in the plane PP', just above the slit, and parallel beams diffracted at angles slightly greater or less than the angle of incidence are brought to a focus at various points along the photographic plate.

The spectrum is not normal. When only a short spectral range is under consideration, calculations may nevertheless be made upon the basis of a normal spectrum, employing equation (6). The deviations from normality can be taken care of by plotting a correction curve (see previous footnote). When a longer spectral range is under consideration, wave-lengths may be calculated by an equation which involves only one additional constant, a correction curve being finally plotted in any case.

The spectrographs represented in Figures 123 and 124 can be used for work in the visible, for work in the ultra-violet to the limit of transmission of the lenses, and for work in the infra-red to the limit to which photographic plates can be sensitized, which is about 12,000 Å (1.2 μ).

21. The Infra-red Spectrometer.—Figure 125 represents a spectrometer for work in the infra-red beyond the photographic limit. An image of the source, S, is formed by the mirror M_1 upon the slit S_1. The beam which passes through this slit falls upon the mirror M_2 and is reflected as a parallel beam to the grating G. The parallel diffracted beam leaving the grating,

* In order to minimize the errors due to slight non-uniformity in the dispersion, the positions of numerous lines in the known spectrum, in addition to λ_1 and λ_2, are always determined. The wave-lengths are then calculated by eq. (6) and the results so obtained are compared with the already known wave-lengths of the lines in question. From the differences a correction curve is plotted; the corrections obtained from this curve are thereupon applied to the wave-lengths of the unknown lines as calculated by eq. (6).

which is indicated by the broken lines, falls upon the mirror M_2 where it is reflected as a converging beam toward the plane mirror M_3, whence it is reflected to the thermopile T, which is connected to a galvanometer. The deflection of the galvanometer is proportional to the intensity of the radiation falling on the thermopile. The directions of the incident and diffracted beams remain fixed. The spectrum is explored by turning the grating by small steps, steps of the order of a minute of arc. The set-

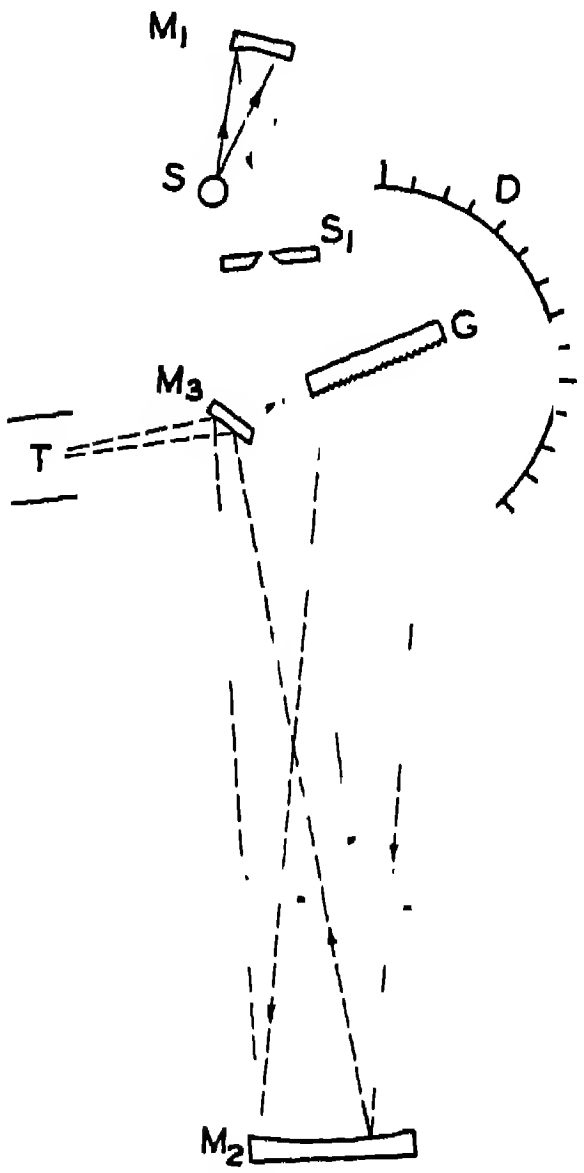


FIG. 125

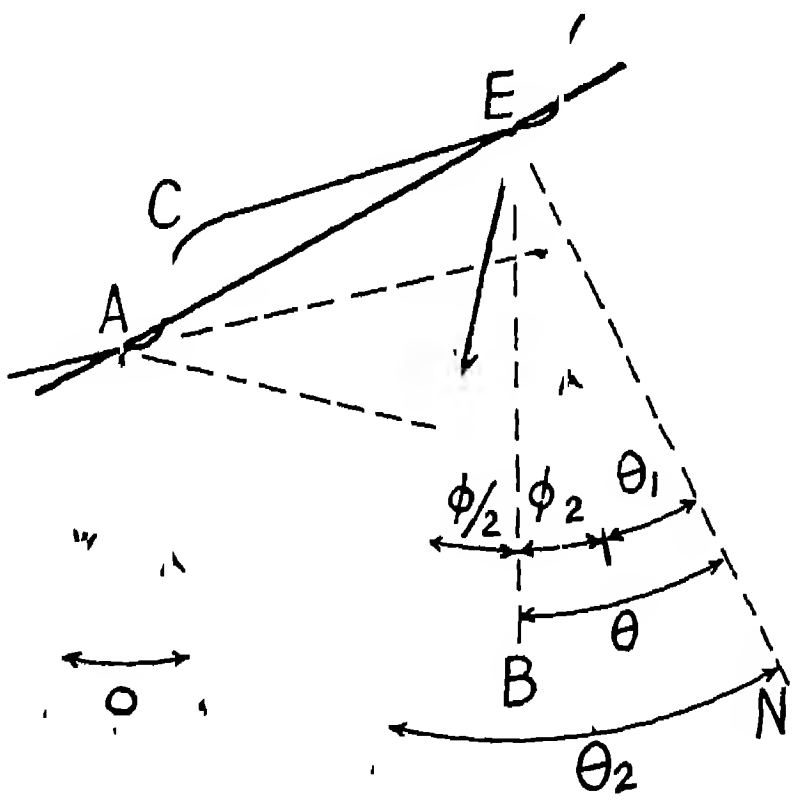


FIG. 126

ting of the grating is read on the divided circle D. Wave-lengths are determined from the grating equation, which is used in a modified form. This form will be derived and some general information will be given:

Referring to Figure 126, suppose A and E are corresponding points in adjacent grooves of the grating, the groove form being represented by the curved line ACE. The straight line joining A and E is parallel to the face of the grating and EN represents the normal to this face. The angle θ_1 represents the angle of incidence and θ_2 the angle of diffraction. The line EB represents the bisector of the angle between the incident ray and the diffracted ray, and θ represents the angle which this bisector makes with normal EN. Denoting the angle between the incident and the diffracted ray by ϕ, we have $\theta_1=\theta-(\phi/2)$ and $\theta_2=\theta+(\phi/2)$. Inserting these values in the grating equation and then applying well-known trigonometric formulae:

$$m\lambda=a\left[\sin\left(\theta-\frac{\phi}{2}\right)+\sin\left(\theta+\frac{\phi}{2}\right)\right]$$

$$=2a\cos\frac{\phi}{2}\sin\theta$$

or

$$m\lambda=K\sin\theta, \tag{7}$$

where $K = 2a \cos(\phi/2)$ is a constant which is known as the *spectrometer* constant. In order to determine the value of θ which corresponds to a given angular setting of the grating, as read upon the divided circle, the setting for which θ equals zero is determined. At this setting the normal EN coincides with the bisector EB and then the central image falls upon the thermopile. The setting at which this happens is determined experimentally. The spectrometer constant is usually determined by making observations upon a spectrum line of known wave-length but may also be determined by computation from the grating constant and the value of the angle ϕ.

In the spectrometer which is represented in Figure 125, the incident beam and the diffracted beam each make a small angle with the axis of the principal mirror M_2, and the result is that the diffracted image of the slit S_1 which is formed upon the slit of the thermopile is not as perfect as it would be if these beams traveled parallel to the axis of the mirror in question. A spectrometer which avoids this difficulty has been described and successfully used by A. H. Pfund,* and spectrometers of the Pfund type or of some other type in which the beams travel strictly parallel to the axis of the principal mirror or mirrors will undoubtedly be extensively used in the future.

To avoid the overlapping of spectral orders which, as we have learned, is a common feature of all grating spectra, a so-called "foreprism spectrometer" or "monochromator" may be used in conjunction with the grating spectrometer. Radiation from the source is allowed to enter the monochromator which produces a spectrum of low dispersion of which a limited region in the infra-red enters the slit of the grating spectrometer. This region is not narrowly limited but comprises a rather wide range of wave-lengths which is selected by setting the monochromator. The function of the monochromator is to select energy of a given range of wave-length and to inhibit the passage of energy of one-half this wave-length, or one-third, or one-fourth, etc. It is not usually necessary to use the monochromator when investigating emission spectra, but it is necessary to use it when investigating absorption spectra, particularly when investigating the band spectra of gases.

22. The Echelette.—Gratings for use in the infra-red are ruled with a relatively large grating space because of the relatively great wave-length of the radiation which is to be investigated. When the grating space is large the grooves must be correspondingly deep, and even a diamond will not stand up under the stress of ruling deep grooves in a hard metal such as speculum. Accordingly, coarse gratings are ruled in comparatively soft metal. Copper, aluminum, tin, and solder have been used with success.

R. W. Wood, who, with A. Trowbridge, explored the effectiveness of coarse gratings with flat grooves, has given the name "echelettes" to such gratings.† Wood, at first, ruled the gratings with a carborundum crystal.

* *JOSA and RSI*, **14**, 337, 1927; see also J. D. Hardy, *Phys. Rev.*, **38**, 2162, 1931.

† *Phil. Mag.*, **20**, 770, 1910; Trowbridge and Wood, *ibid.*, p. 886; also Trowbridge and Crandall, *ibid.*, **22**, 534, 1911; or, summarized account in Wood, *Physical Optics* (Macmillan, 1914), pp. 227–31.

They are now ruled with a diamond which has been ground, and it is attempted not only to have the sides of the grooves flat, but also to have them meeting perpendicularly (see Fig. 127). This form of groove can be reasonably well attained in coarse gratings.

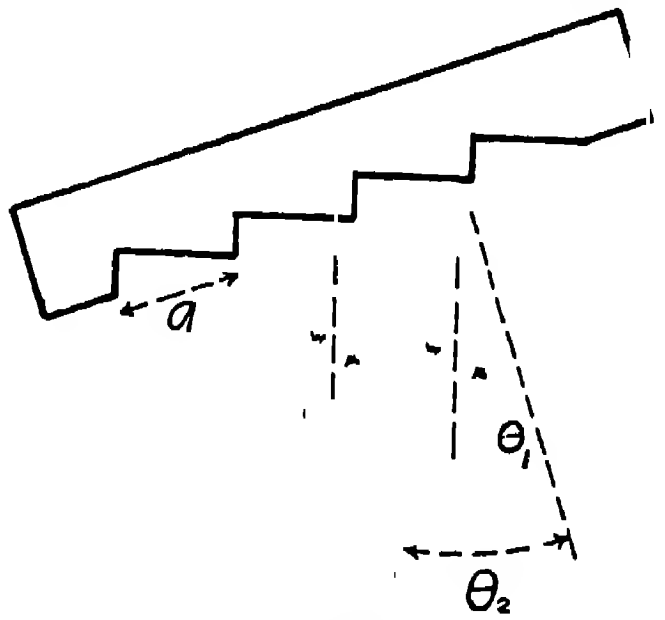


FIG. 127

The flat face of one groove acting alone reflects a maximum of energy into the direction for which the angle of reflection from this face equals the angle of incidence upon it. The spectral order which lies in this general direction will be stronger than other orders. A selective effect is thus obtained. When the angles of incidence and diffraction are varied by rotating the grating, as in the infra-red spectrometer which has been described, a given grating yields satisfactory intensity over a range of rotational angle of 10° or more. The selective effect is not a sharp one, and it is not to be expected that it should be sharp. The reflecting groove face constitutes the equivalent of a slit having the width of the face. This width is as a rule only one or two wave-lengths of the radiation under investigation, and we know that so narrow a slit must scatter energy over a wide range of angle. Consequently, it is unavoidable that the grating should waste some of the available energy in the central image, and, depending upon conditions, perhaps also in spectral orders other than the one under examination.

When visible light falls upon an echelette grating much shorter wave-lengths are involved and each groove face then reflects more nearly according to the laws of geometrical optics. The central image may then be practically absent. When one views the image of one's face in an echelette, one sees the reflection not on the normal to the *grating* as ordinarily, but off to one side, on the normal to the *groove face*. There are also other interesting effects. For an account of these the above-cited references should be consulted.

There is no sharp line of demarcation to be drawn between what constitutes an ordinary grating and what constitutes an echelette. Speculum gratings of 15,000 or even 20,000 lines per inch are used for investigating the infra-red adjacent to the visible and extending to nearly 2μ. In passing thence to regions of greater wave-length, it is advantageous to use successively coarser gratings and there is no abrupt transition in type.

THEORY OF THE CONCAVE GRATING
(Secs. 23–27)

Rowland discovered in the year 1881 that a grating ruled upon a concave surface focuses the spectra which are formed, without the aid of lenses or mirrors.* The concave grating is of value for three principal reasons:

1. It saves the very considerable cost of the large and highly perfect lenses which are necessary with plane gratings.

2. It focuses overlapping spectral orders accurately along the same focal curve. This is a matter of importance when an unknown spectrum is photographed in a given order and a comparison spectrum is photographed in a different order, as is often done. When plane gratings are used in conjunction with lenses, overlapping spectra of different orders are never quite in focus at the same time, because of the residual chromatic aberration of the lenses.

3. The concave grating renders inestimable service in investigating the ultra-violet, especially the extreme ultra-violet region, where the opacity of materials largely precludes the use of lenses for collimating and focusing. The use of mirrors for this purpose is also impracticable because the reflecting power of all substances is extremely low throughout this region.

23. General Properties.—Referring to Figure 128, which we shall consider as lying in a horizontal plane, the grating is represented in central cross-section at G. The lines or grooves are perpendicular to the plane of the figure, or vertical. Denoting the radius of curvature of the grating surface by R, a reference circle of radius $R/2$, represented in broken line, is drawn tangent to the grating at its center. The *diameter* of this so-called "Rowland circle" is equal to the radius of curvature of the grating. The center of curvature of the grating thus lies on the reference circle diametrically opposite the grating. The grating surface lies slightly outside of the reference circle, not along it, except at the central point where the two circles are tangent.

Light from the source S passes through the slit, which lies on the reference circle, and is incident upon the grating. The slit may, theoretically at least, lie at any point on the circle. We shall take up the theory later. Practically, positions of the slit close to the grating are to be avoided—except in one special type of mounting which will be discussed in its proper place. Designating the angle of incidence at the center of the grating by θ_1, and constructing a line on the other side of the normal at an angle equal to θ_1, this line leads to the central image which is formed on the reference circle

* The discovery was announced in several journals during the following year, e.g., *Phil. Mag.* (5) **13**, 469, 1882, or see *Rowland's Physical Papers*, pp. 487 and 492.

at P_0. A given spectrum line, of wave-length λ and spectral order m, is focused by the grating on the circle in a direction determined by applying the grating law at the center of the grating. Thus a given line in the violet may be focused at V_1 to the left of P_0 and at V_1' to the right of P_0, and a given line in the yellow may be focused at Y_1 and Y_1'. The dispersion in the vicini-

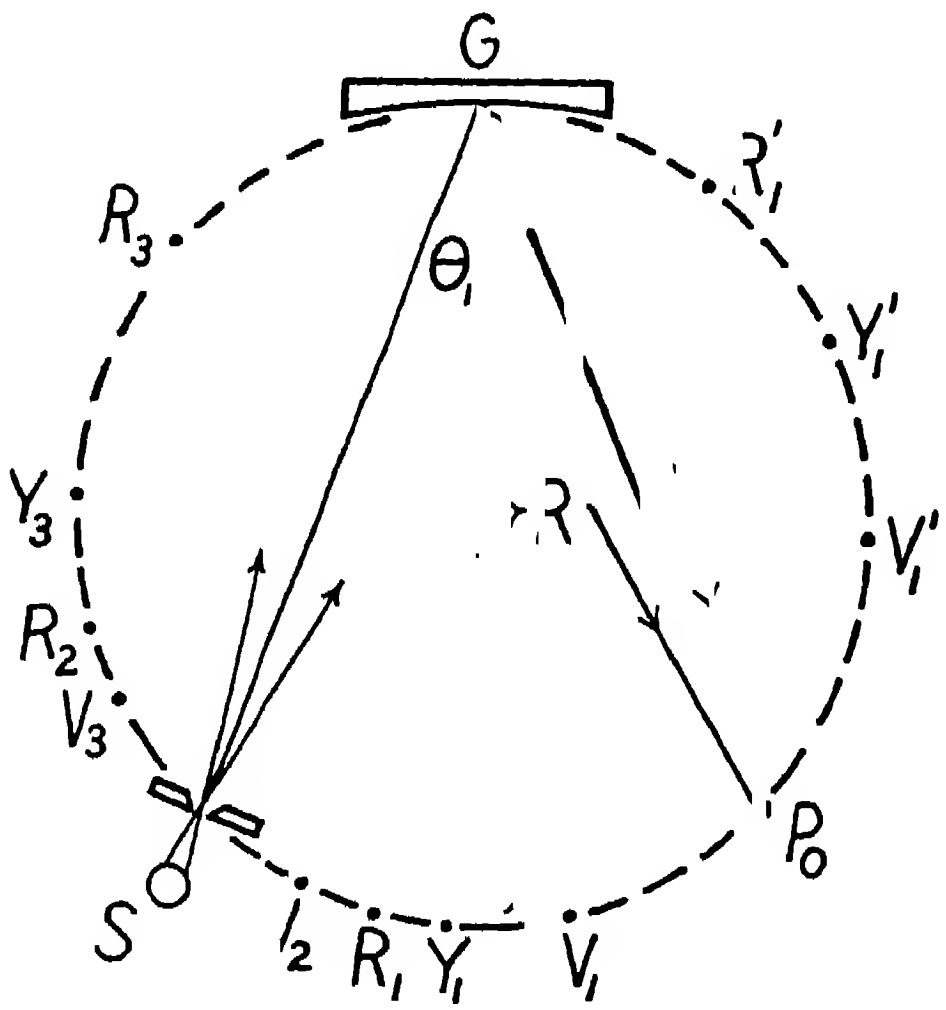


FIG. 128

ty of Y_1, near the normal to the grating, is practically normal. A given spectrum line in the red may be focused at R_1 and R_1'. All of the aforementioned images are supposedly of the first spectral order, and it may be supposed that there will be no possibility of the formation of second-order images in the visible spectrum between R_1' and the grating, the dispersion being too great in this region. However, a second-order image of our violet line will be formed to the left of R_1—let us say at V_2. The second-order image of our yellow line may perhaps coincide in position with the slit, the diffracted beam in this case doubling back on the incident beam. The second-order image of the red line may fall at R_2. The red of the second order overlaps with the violet of the third order, which is represented at V_3. Positions of the third-order yellow and red are indicated at Y_3 and R_3. In this region there will be overlapping with a considerable portion of the fourth-order spectrum and even with the violet of the fifth order.

When the slit is moved clockwise along the Rowland circle, the spectral images move counter-clockwise along this circle.

The concave grating is ruled on a *spherical* surface and not on a cylinder or on a parabola. The radius of curvature of Rowland's large gratings is

usually 21 feet, or 640 cm, and the width of the ruled portion is commonly about 5 inches—or, let us say, 12.8 cm. The angular aperture is thus about $12.8/640 = .02$ radians. Small concave gratings are often ruled with more than proportionately smaller radii of curvature, and these may subtend apertures as high as .08 radians. Other things being equal, the difficulty of ruling increases with the angular aperture subtended. The concavity of the surface of a large grating is so slight that it is likely to escape notice at a cursory glance. The depth of the concavity, the distance DG of Figure 129, may be evaluated by the sagittal relation. For a large grating this depth is $DG = AD^2/2R = (6.4)^2/(2 \times 640) = .032$ cm.

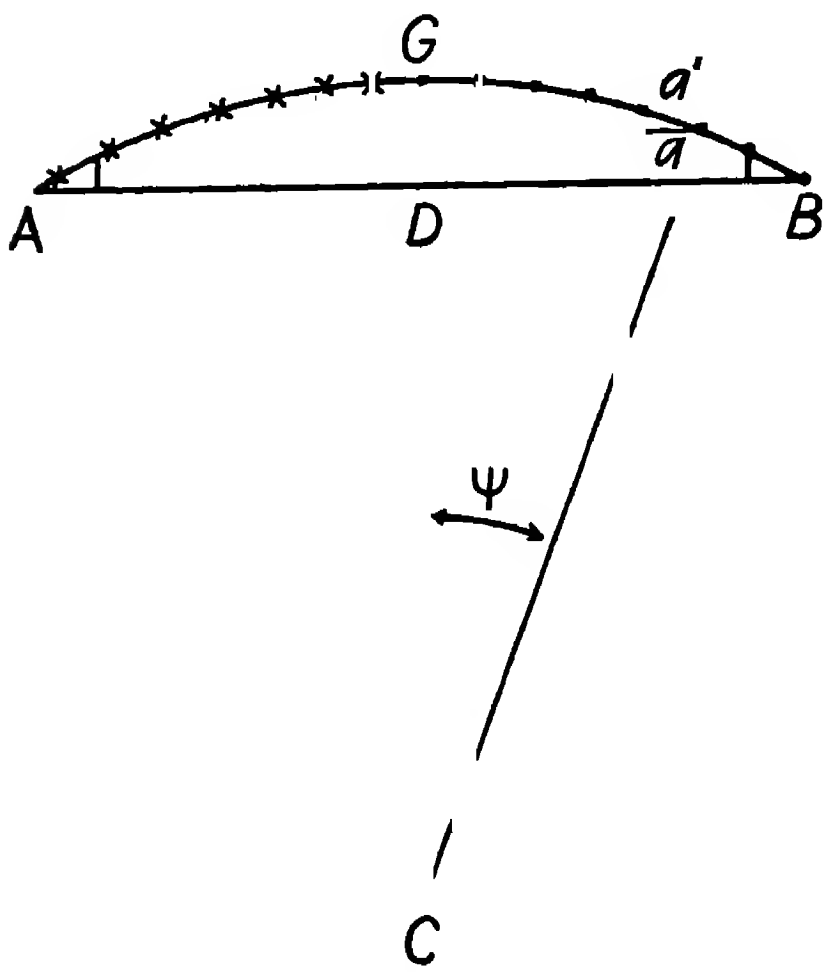


FIG. 129

Theory requires, as we shall learn, that the lines of a concave grating should be equally spaced along a *chord* drawn across the surface and *not* equally spaced along the arc of the surface. A concave grating is ruled in much the same manner as a plane grating—this manner of ruling obviously causes the lines to be equally spaced along a chord as they should be. In Figure 129 the arc AGB represents the grating surface and the straight line AB represents the chord. Perpendiculars erected at equal intervals along the chord strike the grating surface at corresponding points of adjacent lines or grooves of the grating. These positions are indicated by dots along the right-hand portion of the arc. Along the left-hand portion of the arc small crosses indicate what *would be* corresponding points of adjacent grooves if the lines were equally spaced along the arc. In passing from the center outward the crosses fall progressively farther behind the corresponding perpendiculars. There is a *cumulative* difference in position involved.

We shall henceforth, for the sake of brevity, instead of speaking of "the positions of corresponding points of adjacent lines or grooves" simply speak of "the positions of adjacent lines." This is merely for purpose of abbreviation and does not imply that the grooves are narrow compared to the intervening unruled portion of the surface. The grooves practically fill the ruled surface.

Since the lines are equally spaced along the chord, there is a *very* slight

increase in the spacing measured along the arc in passing from the center of the grating toward either edge. Denoting the variable grating space along the arc by a', and the constant grating space along the chord by a, we have:

$$a' = \frac{a}{\cos \psi}, \tag{8}$$

where ψ is the angle between the chord and the arc at the point in question, or the equal angle indicated in the figure between the normals erected respectively at the center of the grating and at the point in question.

The grating space along the arc has a *minimum* value of $a' = a$, at the center of the grating, but even at the edge of the grating, a' is greater than a by only one part in twenty-thousand, as will develop later.

24. Theory of Focusing along Rowland Circle.—The theory of the concave grating is usually given following the treatment of C. Runge. This treatment is complete from a mathematical standpoint but fails to impart a clear physical conception of the cause of the focusing. The Runge treatment is accessible in numerous places.* It would be superfluous to repeat it here. In its stead a treatment will be given which emphasizes the physical aspects of the theory.

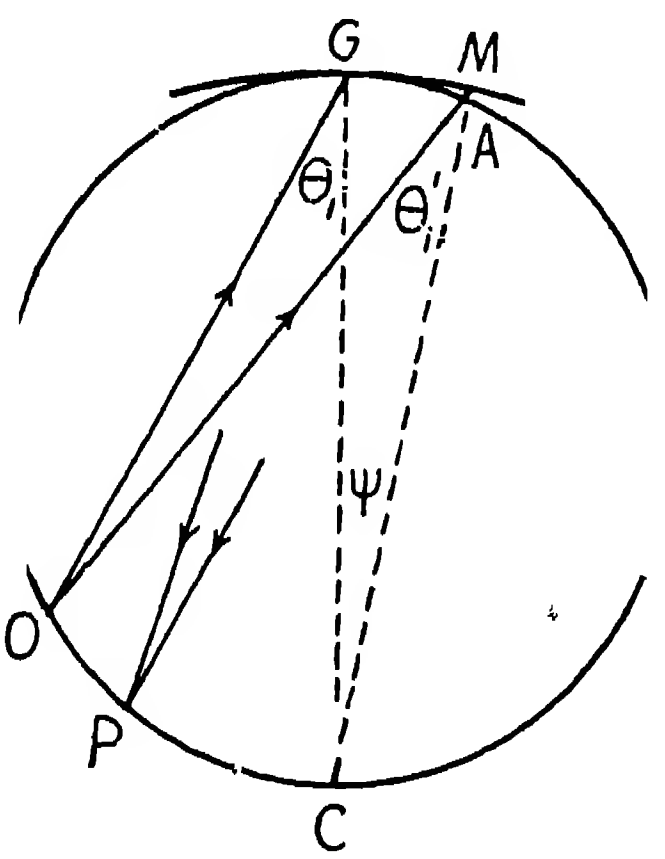


FIG. 130

Referring to Figure 130, let us denote the angle of incidence at the center of the grating by θ_1 and the angle of incidence at any other point M of the surface by θ_1'. The angle θ_1' varies as the position of M varies, and this angle passes through a *maximum* value of $\theta_1' = \theta_1$ at the center, as is evident from the following considerations: At various points lying on a circle —for example, at points G and A—a given arc—for example, OC—subtends equal angles. Any point M of the grating not coinciding with G lies outside of the circle $OGAC$, the Rowland circle, and hence the angle θ_1', subtended at any point M by the arc OC, is less than the angle θ_1 subtended at the center of the grating by the arc OC. Hence θ_1' passes through a maximum value of $\theta_1' = \theta_1$ at the center. In the same way the angle of diffraction, θ_2', not indi-

* Runge never published the treatment himself. It was first published in abstract by H. Kayser in Winkelmann, *Handb.* (1st ed.), **2**, Part I, 407, and later in full by Kayser in his *Handb. d. Spektroskopie*, **1**, 452. Abstracts are to be found also, e.g., in Wood, *Physical Optics*, p. 233, and Schuster and Nicholson, *The Theory of Optics* (3d ed.), p. 126.

cated in the figure, passes through a maximum value of $\theta_2' = \theta_2$ at the center. The variations of θ_1' and θ_2' are *very* slight.

We now apply the grating law at every element of the grating surface: The path from O by way of one line of the grating to P differs from the path by way of the next adjacent line by an amount δ, which is determined by the equation $\delta = a' \sin \theta_1' + a' \sin \theta_2'$, where a' is the grating space measured along the arc. The element of the grating surface is to be thought of as being wide enough to include numerous lines—a millimeter wide, let us say. The mathematical condition that P shall be a diffracted image of O is that as the grating is traversed δ shall remain constant. If δ is constant there will be a spectral image of order m and wave-length λ formed according to the grating law, $\delta = m\lambda = a' \sin \theta_1' + a' \sin \theta_2'$. From a physical standpoint we may say that an image will be formed if the variation of δ does not exceed some extremely small permitted value.

The grating space along the arc, a', passes through a minimum value at the center of the grating, hence the differential of a' is there equal to zero, that is, $da' = 0$. When O and P lie on the Rowland circle, θ_1' and θ_2' pass through maximum values at the center, hence $d \sin \theta_1' = 0$ and $d \sin \theta_2' = 0$ at this point. Therefore, in the vicinity of the center of the grating:

$$d\delta = d(a' \sin \theta_1' + a' \sin \theta_2') = 0 \tag{9}$$

and hence in the vicinity of the center:

$$\delta = m\lambda = a' \sin \theta_1' + a' \sin \theta_2' = \text{Const.} \tag{10}$$

This equation, interpreted physically, tells us that when O and P both lie on the reference circle, the condition for the formation of a diffracted image is satisfied for at least a certain width on either side of the center of the grating. How wide the grating may be will depend upon how nearly the decrease in $\sin \theta_1'$ and $\sin \theta_2'$ which occurs in passing from the center of the grating outward is compensated for by the increase in a' which occurs in passing from the center outward. We shall learn that the compensation is complete to within a degree of approximation used in evaluating one small term which is involved in the proof.

Consider the path difference $\delta = a' \sin \theta_1' + a' \sin \theta_2'$ divided into two parts, $\delta_1 = a' \sin \theta_1'$ and $\delta_2 = a' \sin \theta_2'$. We shall show that the path difference δ_1 remains constant in traversing the grating, and we shall then be able to infer that δ_2 also remains constant because δ_1 and δ_2 play similar rôles in the theory.

For an element of the grating surface chosen at the center $\delta_1 = a \sin \theta_1$. If δ_1

is to have this same value for a lateral element chosen at M (Fig. 130), we must have $a' \sin \theta_1' = a \sin \theta_1$, or:

$$\frac{a'}{a} = \frac{\sin \theta_1}{\sin \theta_1'}. \tag{11}$$

We must therefore show that this equation holds for a grating ruled with constant spacing along the chord. Introducing the value of a'/a obtained from equation (8) and referring to Figure 131, where the angle OAC has a value equal to θ_1 and the angle OMC is θ_1', we have:

$$\frac{a'}{a} = \frac{1}{\cos \psi} = \frac{CG}{CA} = \frac{CM}{CA} = \frac{CA+AM}{CA} = 1 + \frac{AM}{CA}. \tag{12}$$

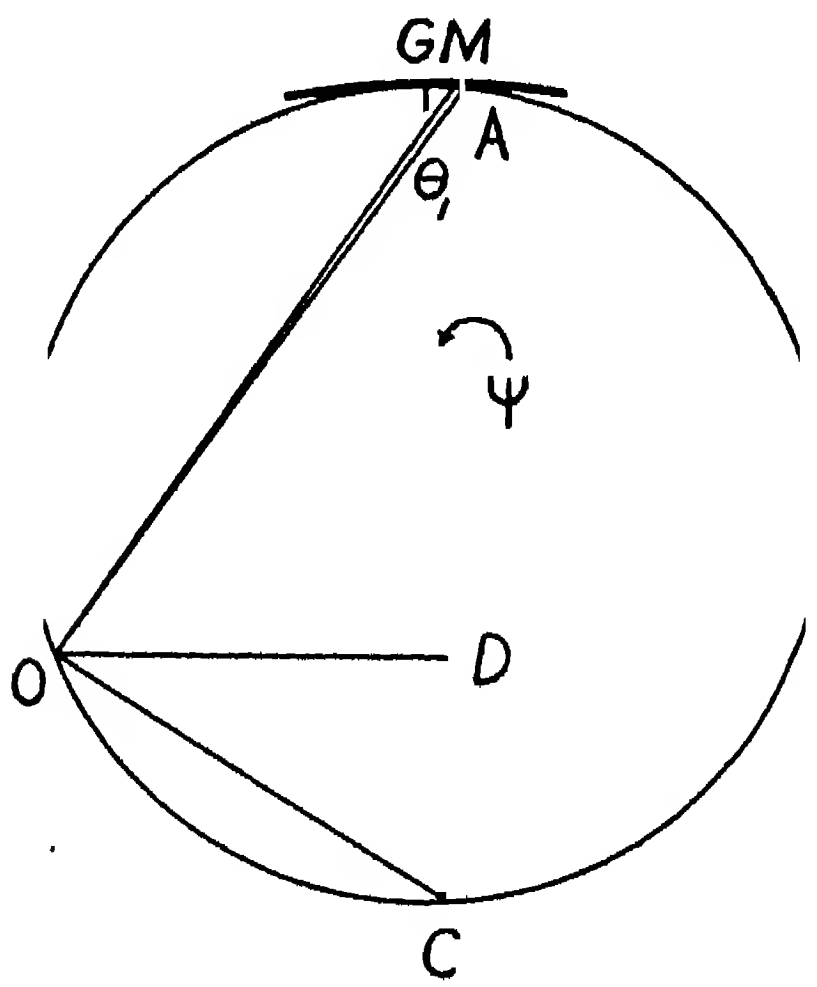


FIG. 131

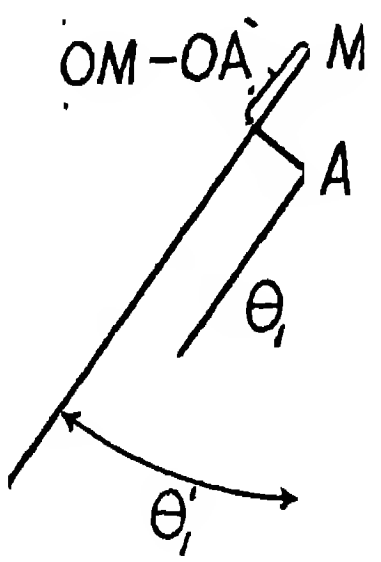


FIG. 132

This equation is exact—there is on approximation involved. Proceeding now to evaluate $\sin \theta_1 / \sin \theta_1'$, we have $\sin \theta_1 = OD/OA$ and $\sin \theta_1' = OD/OM$. Hence:

$$\frac{\sin \theta_1}{\sin \theta_1'} = \frac{OM}{OA} = \frac{OA + (OM - OA)}{OA} = 1 + \frac{(OM - OA)}{OA}. \tag{13}$$

The distance represented by $(OM - OA)$ is indicated at the top of Figure 132 to a large scale. Now $(OM - OA) = AM \cos \theta_1' = AM \cos \theta_1$, to a degree of approximation which is ample for evaluating the final term of equation (13), which is small and plays only the rôle of a "correction term." Moreover,

$OA = CA \cos \theta_1$ approximately, when M lies close to G. The distance from M to G is much exaggerated in our figure, where the width of the entire grating if drawn to scale would be only about a millimeter. According to the foregoing approximations, $(OM - OA)/OA = AM \cos \theta_1 / CA \cos \theta_1 = AM/CA$. Substituting this value in the right-hand side of equation (13), the right-hand sides of equations (12) and (13) become identical. Hence the left-hand sides of these equations are equal. In other words, equation (11) is verified, and from this it follows that $\delta_1 = a' \sin \theta_1'$ is constant. The constancy of $\delta_2 = a' \sin \theta_2'$ may be demonstrated in the same manner, and hence we conclude finally that $\delta = \delta_1 + \delta_2$ is constant across the grating.

We have now shown that when the grating is ruled with constant spacing along the chord, the variation in the spacing a', along the arc, compensates for the variations in the angles of incidence and diffraction which occur across the grating, provided the source and field point lie on the Rowland circle. Thus, when the slit lies on this circle, diffracted spectral images are formed on this circle. However, our proof has involved an approximate evaluation of one small term, and to be quite sure of our ground we must convince ourselves that this approximation is *sufficiently* close. The permissible error in the approximation is evaluated toward the end of the next section, and it will become evident that, actually, this is not exceeded.

It is possible to obtain a *higher* approximation than the one which we have used. With the aid of this higher approximation the "aberration" of the concave grating may be deduced. This is small under ordinary circumstances. The calculation is carried out in Appendix D, the reading of which should, however, be deferred until the next section has been read.

25. Some Numerical Evaluations.—The theory of the concave grating can be thoroughly appreciated only when the magnitudes of the quantities involved are understood. Accordingly, we shall evaluate some of these, taking as an example a large Rowland grating.

Let us begin by determining the value of the grating space along the arc, a', in terms of the constant grating space, a, along the chord, at a point halfway out to the edge of the grating, and again at the edge. According to equation (8), $a' = a/\cos \psi$. The development of $\cos \psi$ in series is $\cos \psi = 1 - \psi^2/2 + \psi^4/\underline{4}. \ldots$ Since ψ is small, we may write $1/\cos \psi = 1 + \psi^2/2$, neglecting higher terms. If the angular aperture of the grating is .02 radian, then:

At a point halfway out to the edge:

$$\psi = .005 \text{ radian, and } a' = a\left(1 + \frac{\psi^2}{2}\right) = a\left(1 + \frac{1}{80{,}000}\right)$$

and:

At the edge:

$$\psi = .01 \text{ radian, and } a' = a\left(1 + \frac{1}{20{,}000}\right).$$

Thus the variation in a' is very small, but it has nevertheless a decided significance.

If the small variation in the grating space did not occur, then the positions of the lines would fall behind their actual positions in the manner illustrated in the left-hand portion of Figure 129. At a given point of the surface the position difference in question equals the difference between the length of the arc from the center of the grating to the given point and the length of the semichord to this point. The length of the arc is $R\psi$, where R is the radius of curvature, and the length of the semichord is $R \sin \psi$. Developing $\sin \psi$ in series, $\sin \psi = \psi - \psi^3/\underline{3} + \psi^5/\underline{5} \ldots$. We need retain only the first two terms of this series, and hence:

$$\text{Position difference} = R\psi - R \sin \psi = \frac{R\psi^3}{6}. \tag{14}$$

At a point halfway out to the edge of the grating, this yields $[640 \times (.005)^3]/6 = .13 \times 10^{-4}$ cm, which is about one-tenth of a grating space for a grating having 20,000 lines per inch (for which $a = 1.27 \times 10^{-4}$ cm). If the remainder of the grating were blended off on each side, the position difference at the edge of the effective portion of the grating would thus be about one-tenth of a grating space. A position error of this magnitude at the edge of a grating may be tolerated. Therefore, if we were to content ourselves with a grating of half the width of the actual one, the grating would render about equally good service whether the variation in a' occurred or not. But if we wish to use a grating of the full actual width, the situation is quite different.

A position error of one-half a grating space will cause the disturbances to arrive at the diffracted image in just the reverse of the phase in which they should arrive, in the first-order spectrum. In the second-order spectrum an error of one-quarter of a grating space will produce the same result. The error of optical path would in each case be $\lambda/2$, and such an error should obviously not occur. However, a path error of $\lambda/4$ for disturbances coming from the edge of the grating might be tolerated—no great loss of definition or of intensity would result from an error of this magnitude.* A path error of $\lambda/4$ is produced in the nth-order spectrum when the position error is the

* For further details see sec. 1 of chap. 7, on "Permitted Aberration."

$1/4m$ part of a grating space, or $a/4m$ cm. If we wish to work in the third-order spectrum, a position error of one-twelfth, or .083 of a grating space, may just be tolerated.

Let us now see how great the position difference at the edge of the actual grating would be if the variation in a' did not occur. At the edge of the grating $\psi = .01$ radian. According to equation (14), the position difference is $[640 \times (.01)^3]/6 = 1.07 \times 10^{-4}$ cm, or about .84 grating space. This position difference of .84 grating space equals the value of the position error which would obtain at the edge of the grating if the variations in the angles of incidence and diffraction which occur across the grating were not compensated by a corresponding variation of the grating space along the arc. This error is about ten times the error, .083, which we have learned is permissible for work as high as the third order.

Our previous proof indicated that the variations in the angles of incidence and of diffraction are exactly compensated when the grating is ruled with equal spaces along the chord. However, the proof involved an approximation. Now we see, according to the foregoing paragraph, that if there were no compensation at all, the resulting error would be ten times the permissible error. Accordingly, the *variations* in the sines of the angles of incidence and of diffraction need to be compensated to the extent of 90 per cent, but only to this extent. In other words, as much as 10 per cent of the total variation, in itself extremely small, does not need to be compensated. Thus the approximation which we made in our proof might permissibly be in error by as much as 10 per cent. Evidently it is not in error by as much as this. Therefore the compensation is *sufficient*—the aberration does not exceed the permitted limit.*

The aberration is evaluated in Appendix D.

It is a fortunate coincidence that theory requires the lines of a concave grating to be equally spaced along the chord since this is the manner in which the machine naturally rules them. However, if theory required that the lines should be spaced otherwise, they could be so ruled. In the process of ruling, the long nut which fits over the main screw of the ruling machine would be given a slight rotation as it advances along the screw. A very slight rotation would suffice. By way of example let us suppose that it were required to space the lines equally along the arc. The requisite rotation of the nut, in ruling the entire grating, would then be less than one one-thousandth of a turn, or less than the angle subtended by one tooth of the spac-

* The approximation which was made in the previous proof is very close when the angular aperture subtended by the grating is small, which it always is, and so long as neither the incident nor the diffracted beam nearly grazes the grating. When one or both of the beams nearly graze the grating the aberration becomes appreciable.

ing wheel at the center of this wheel. That this is true follows from the fact that the correction of the positions of the lines, even at the edge of the grating, would amount to only .84 grating space, as we have learned.

26. General Focal Relation.—Referring to Figure 133, suppose the arc G represents the surface of the grating, C the center of curvature, and O a point source of light lying anywhere in the plane of the figure. We shall seek an equation which relates the position of the diffracted image of order m and wave-length λ, indicated by the point P, with the position of the source.

We shall apply the grating law to every element of the grating surface and in doing this shall introduce the grating space as a constant. It may seem upon first thought that this procedure will limit our result to gratings having constant spacing along the arc and thus exclude concave gratings as actually ruled, with constant spacing along the chord. This limitation does not apply, however. If we were to discuss aberration—or, in other words, discuss the perfection of the image—the objection would be well founded. But we shall limit ourselves to discussing the *position* of the image, and this position will be the same whether the spacing be constant along the chord or constant along the arc. For a sufficiently narrow grating the two types of spacing are physically indistinguishable, and it is obvious that the position of the image must be the same for a narrow grating as it is for a wide one of which the narrow grating may be considered to form a part.

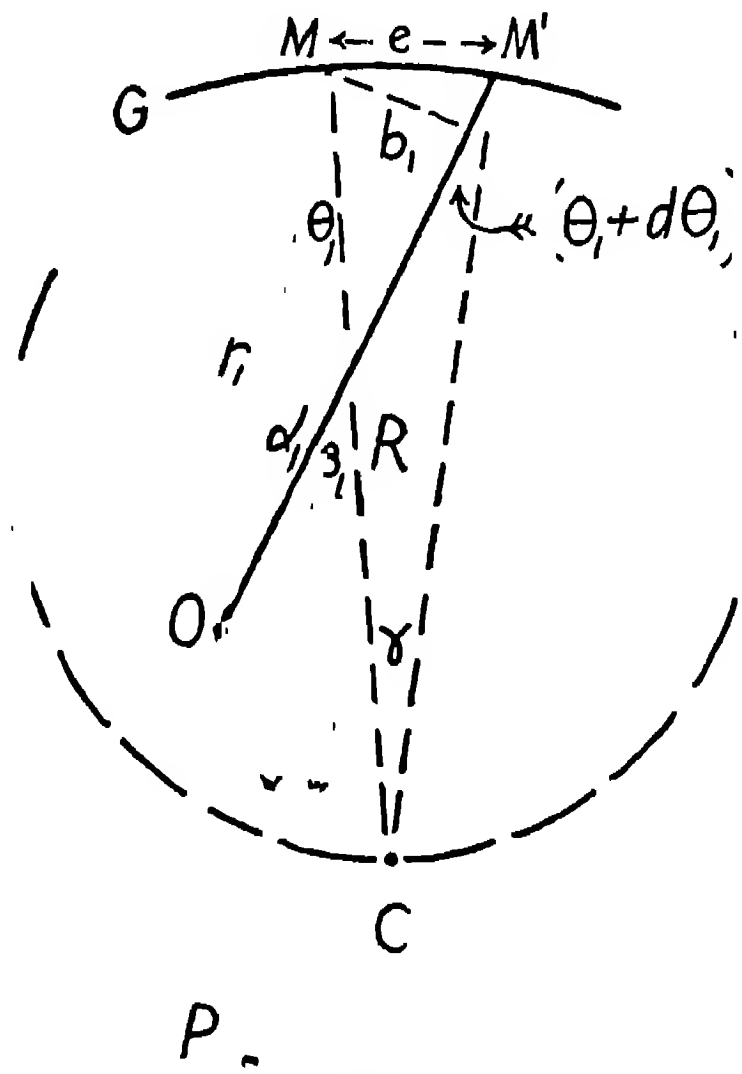


FIG. 133

Denote the angle of incidence at any point M of the grating surface by θ_1, as indicated in the figure, and the angle of diffraction by θ_2, not indicated. The angles θ_1 and θ_2 both vary across the grating, and, moreover, since O and P do not, in general, lie on the reference circle, these angles may now vary by considerable amounts.

The grating equation must apply at every element of the grating surface. Differentiating this equation $d(m\lambda)=d[a(\sin\theta_1+\sin\theta_2)]=0$. Since the grating space a is to be regarded as a constant, we have:

$$d(\sin\theta_1+\sin\theta_2)=0$$

or

$$\cos\theta_1 d\theta_1 + \cos\theta_2 d\theta_2 = 0\,. \tag{15}$$

If an image is to be formed, the variations in θ_1 and θ_2 must be such as to compensate for each other, according to the foregoing equation.

Let M' be a point on the grating surface which lies at a *small* distance e from M. Denote the radius of curvature of the grating by R and the distance OM by r_1. We shall derive an expression for $d\theta_1$ in terms of e, R, θ_1, and r_1, and shall then be able to write by analogy a similar expression for $d\theta_2$ in which r_2 will be used to denote the distance PM. We have the following relation between angles indicated in the figure:

$$\alpha_1 + \theta_1 = \beta_1 = \gamma + (\theta_1 + d\theta_1)\,,$$

whence:

$$\begin{aligned} d\theta_1 &= \alpha_1 - \gamma = \frac{b_1}{r_1} - \frac{e}{R} = \frac{e\cos\theta_1}{r_1} - \frac{e}{R}\,, \\ d\theta_2 &= \frac{e\cos\theta_2}{r_2} - \frac{e}{R}\,. \end{aligned} \tag{16}$$

Substituting these equations in equation (15) and dropping the common factor e, we have:

$$\cos\theta_1 \left[\frac{\cos\theta_1}{r_1} - \frac{1}{R}\right] + \cos\theta_2 \left[\frac{\cos\theta_2}{r_2} - \frac{1}{R}\right] = 0\,. \tag{17}$$

This is the desired equation. The grating law is to be invoked in conjunction with it when necessary.

This equation is satisfied when the source and field point both lie on the Rowland circle. Then, referring to Figure 134, $r_1 = R\cos\theta_1$, or $[(\cos\theta_1)/r_1 - 1/R] = 0$, and, similarly, $r_2 = R\cos\theta_2$ or $[(\cos\theta_2)/r_2 - 1/R] = 0$. Thus we again arrive at the focal relation as given by the Rowland circle.

For a plane grating R has the value ∞, and equation (17) becomes:

$$\frac{\cos^2\theta_1}{r_1} + \frac{\cos^2\theta_2}{r_2} = 0\,. \tag{18}$$

This equation enables us to calculate the position of the virtual diffracted image when a divergent beam falls upon a plane grating. This case was discussed in a general way in section 18.

We shall have occasion to make further use of equation (17). This equation was first derived by Rowland, by a much longer process, however, and

in different form. It was independently derived by E. Mascart* in the interval which elapsed between Rowland's first brief announcement of his success with concave gratings and the publication of his later more detailed

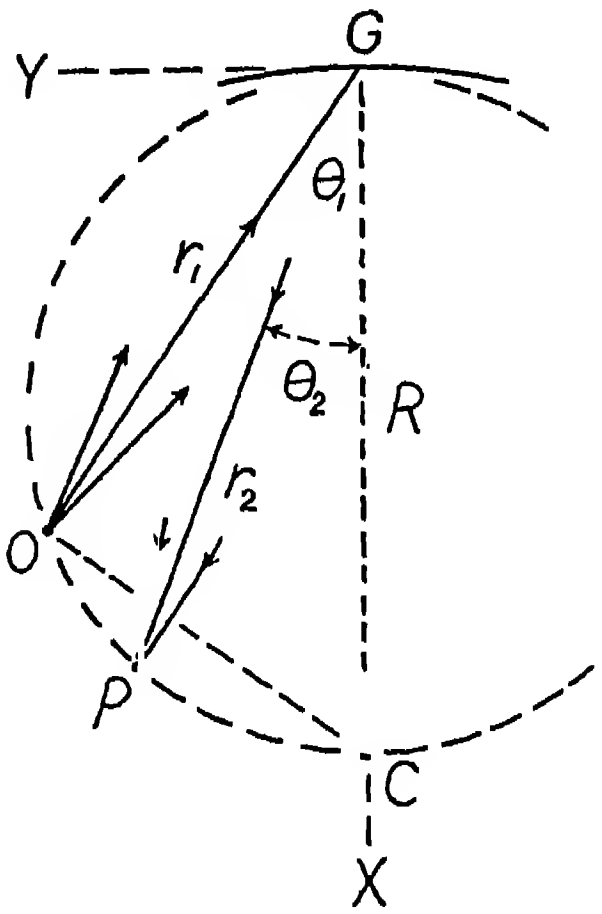


FIG. 134

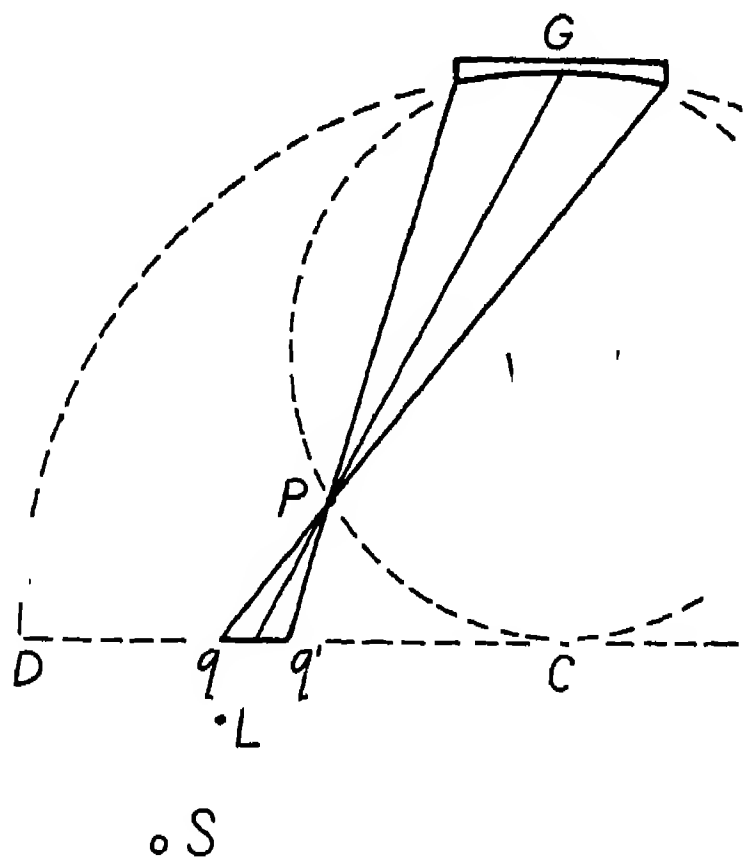


FIG. 135

account.† The equation may be transformed by simple algebraic steps into the following form in which it is frequently given:

$$r_2 = \frac{Rr_1 \cos^2 \theta_2}{r_1(\cos \theta_1 + \cos \theta_2) - R \cos^2 \theta_1}. \tag{19}$$

This form permits a direct evaluation of r_2—the distance from the grating to the point P (Fig. 133) where the image is formed.

27. Astigmatism.—We have thus far considered the focusing of only those rays which lie in the central horizontal plane, or, referring again to Figure 134, of those rays which lie in the plane of the figure. This plane is frequently referred to as the "XY-plane," a system of co-ordinates being taken as indicated, with the center of the grating surface as origin, the X-axis perpendicular to the surface, and the Y-axis tangent to the surface and perpendicular to the direction of the ruling. The Z-axis is then taken parallel to the direction of ruling, out from the plane of our figure, or vertically upward when the grating is set up with the lines vertical as usually supposed. The rays, originating from the source O, which lie in the XY-plane, consti-

* *Jour. d. phys.* (2), **2**, 5, 1883.

† Other treatments of the concave grating which were submitted shortly after Rowland's discovery are W. Baily, *Phil. Mag.* (5), **15**, 183, 1883, and R. T. Glazebrook, *ibid.*, p. 414, and *ibid.* (5), **16**, 377, 1883.

tute what we may call a horizontal *fan*. This fan, after diffraction by the grating in a given wave-length and spectral order, forms another horizontal fan intersecting at P. However, the rays of any *vertical* fan originating from O do *not*, after diffraction, all pass through P, but, as will appear presently, the rays which are incident above the XY-plane pass above P, and those incident below this plane pass below P. That is, it will develop that the rays of any vertical fan all pass through a vertical line through P, a line perpendicular to the plane of the figure. Thus all of the vertical fans from the point O, considered collectively, form a vertical line image through P. In other words, a *point* source gives rise to a *line* image. Such an image is said to be "astigmatic." The images formed by a grating are, in general, astigmatic.

Let us now consider how this astigmatism arises in the simplest case, namely, when a point source of light lies *at the center of curvature* of the grating, C (Fig. 135). The rays from C which strike the grating are each normal to the surface and are consequently also normal to the ruled lines at the point of incidence. This applies to rays striking above or below the XY-plane as well as to those lying in this plane. The rays diffracted at a given point of the grating surface therefore lie in the plane which is perpendicular to the ruled lines at the point in question. This plane erected at any point of the surface always passes through the line CD, which is that diameter, or that axis of the sphere of which the grating surface forms a part, which is perpendicular both to the ruled lines and to the direction CG. The horizontal fan of rays from C, after diffraction, converges to P and then diverges, spreading over the line segment qq', along the axis CD. If now the incident horizontal fan of rays is rotated about CD as an axis, both the incident and the diffracted rays remain at all times perpendicular to the ruled lines, and it obviously follows that the diffracted rays will always pass through the line segment qq'. A given ray of the rotating horizontal fan always passes through a given point of the line segment qq'. Moreover, a given ray of the rotating horizontal fan generates a vertical fan. Consequently, the rays of a given vertical fan intersect at a given point along qq'. When the horizontal fan is rotated about CD as axis, the point of intersection of this fan, originally P, generates what is, strictly speaking, a short segment of a circular arc about CD as axis, through P, perpendicular to the plane of the figure. However, this segment of arc is so short that, practically, it constitutes a vertical line segment through P. The length of the segment, for a given position of P on the Rowland circle, is directly proportional to the length of the ruled lines of the grating, and for a given length of ruled lines is greater for a position of P far removed from C than for one close to C.

The astigmatism which results when the source occupies a *general position* on the Rowland circle is discussed in Appendix E, where, furthermore, several pertinent formulae are developed. It is shown in this Appendix that there are always two focal lines which are mutually perpendicular. One of these focal lines is vertical and passes through the given field point on the Rowland circle; the other is horizontal and lies outside of this circle —it is the line qq' in the case which we have just discussed. The two focal lines are similar to those produced by a concave mirror when the incident beam makes an angle with the axis of the mirror.

The astigmatism of the concave grating carries with it certain consequences which are of importance to the experimentalist:

When a point source of light is used there is a loss of intensity due to the elongation of the image into a line. This loss is, however, at least partially regained when a vertical slit is used as the source. A slit may be regarded as a succession of point sources along a line. When the slit is vertical the astigmatic images due to the several points of the slit overlap, and all of the lost intensity may be regained when it is feasible to use a slit of sufficient length. More will be said on this subject presently.

The slit should be erected accurately parallel to the lines of the grating. Suppose, for example, that the lines of the grating are vertical and that the slit is slightly inclined. Discrete points of the slit will now produce discrete vertical line images, as indicated by the individual lines of Figure 136. But since a slit is really equivalent to a continuous succession of point sources, the resulting images, instead of being separated, will in reality *fill* an area, the area of our figure. Consequently, when the slit is inclined with reference to the rulings, the spectrum lines are broader than they should be and taper off obliquely at the top and bottom. It is thus evident that the slit should be adjusted to parallelism with the ruled lines. FIG. 136

It is to be noted in Figure 136 that the lower end of the vertical line which lies farthest to the right is just opposite the upper end of the line which lies farthest to the left. When it is feasible to illuminate a length of slit great enough so that this condition obtains, then all of the intensity originally lost due to the astigmatism is regained at the mid-height of the spectrum line—this, however, applies only when the slit is parallel to the rulings, for it is only in this case that the images due to the different points of the slit all lie along the same vertical line. The intensity of the spectrum line is then greatest at mid-height and decreases toward top and bottom.

It is often not feasible to illuminate a sufficient length of the slit to regain all of the intensity originally lost due to the astigmatism. The net loss of

intensity which then results is a serious drawback in the investigation of faint spectra. However, the amount of astigmatism varies widely with the conditions under which the grating is being used; moreover, there is one type of concave-grating mounting which is free from astigmatism.

Now as to a different consequence of astigmatism: In a spectrograph which is *stigmatic*, one in which a point in the source gives rise to a point (and not a line), in the image, it is common practice to place a small reflecting "comparison prism" just outside of the slit. This prism serves to reflect light from an auxiliary source, located off to one side, into the spectrograph. A comparison spectrum may thus be placed as a central horizontal strip in a spectrogram, with the unknown spectrum above and below. In a spectrograph which is *astigmatic*, a comparison prism placed just outside of the slit will not separate the two spectra into distinct horizontal strips. However, as first pointed out by J. L. Sirks,* a comparison prism placed at the proper distance from the slit will separate the spectra in the desired manner. Referring again to Figure 135, suppose the primary source of the unknown spectrum is placed at S, and a convex lens at L focuses the light from S on the slit which is placed at P. If a broad but low comparison prism is placed at qq' and a comparison spectrum is reflected in from the side, a sharp separation of the comparison spectrum and the unknown spectrum, in horizontal strips, will occur over a considerable range in the vicinity of C, the center of curvature of the grating. Or, as also pointed out by Sirks, in place of the comparison prism, complementary diaphragms may be used in succession at qq'. When for some reason the use of Sirks's method is inconvenient, the spectra may be separated by using complementary diaphragms in front of the photographic plate.

To sum up, the astigmatism has the following consequences: Generally speaking, it causes a loss of intensity; it necessitates careful adjustment of the slit to parallelism with the grating lines; it necessitates resorting to a modification of the methods which are generally used for separating two spectra which are to be photographed in juxtaposition.

For references to the literature on the astigmatism of the concave grating see the end of Appendix E.

CONCAVE-GRATING MOUNTINGS

(Secs. 28–33)

The various mountings which will be described are all except one based upon the general principle discovered by Rowland that when the source lies upon the reference circle then the spectra will be in focus on this circle. All

* *Astronomy and Astrophysics*, **13**, 763, 1894.

mountings based upon this principle may properly be called "Rowland-circle mountings," but among them only one is commonly called "the Rowland mounting." The principle of the Rowland circle has been applied by various investigators, including Rowland, in a number of specific ways. Each way of applying the general principle carries with it its own optical and mechanical characteristics, and the mountings are given different names when these characteristics differ. Let us consider first the mounting which is simplest in conception.

28.—The Paschen Mounting.*—This mounting is represented in Figure 137. The main portion of the figure is in a horizontal plane. The grating, G, faces a circular metal frame, PP', which constitutes an embodiment of the Rowland circle. This frame plays the rôle of photographic plate-holder. A vertical slit is mounted at some more or less arbitrarily chosen position, O, on the Rowland circle. A convenient construction for the frame PP' is shown in vertical cross-section in the inset in the lower right-hand corner of the figure where the frame proper is indicated by f and f'. Supporting members, mm', are provided outside of the frame proper, at intervals of, let us say, one or two feet along the circle. The frame has a horizontal slot in it which is narrower than the photographic plates which are to be used. A plate, p, is pressed against the outside face of the frame proper over the slot, and bent to the curvature of the circle by means of clamps. Theoretically the outside face of the frame lies along the Rowland circle. In practice it may deviate from this position by one or more centimeters. The reason for this is that the grating may have so-called "error of run" of a type which may cause the best focus to be a little off of the reference circle. The effect of error of run will be discussed in section 40. The frame must be carefully adjusted to the position of best focus at every point along its length. Once this adjustment has been made, a photographic plate may be placed at any point along the frame and thus any portion of the spectrum may be photographed in any of the available orders without altering the adjustments of the spectrograph. This is a great advantage and has led to popularity of the Paschen mounting. As over against this advantage there are some minor disadvantages.

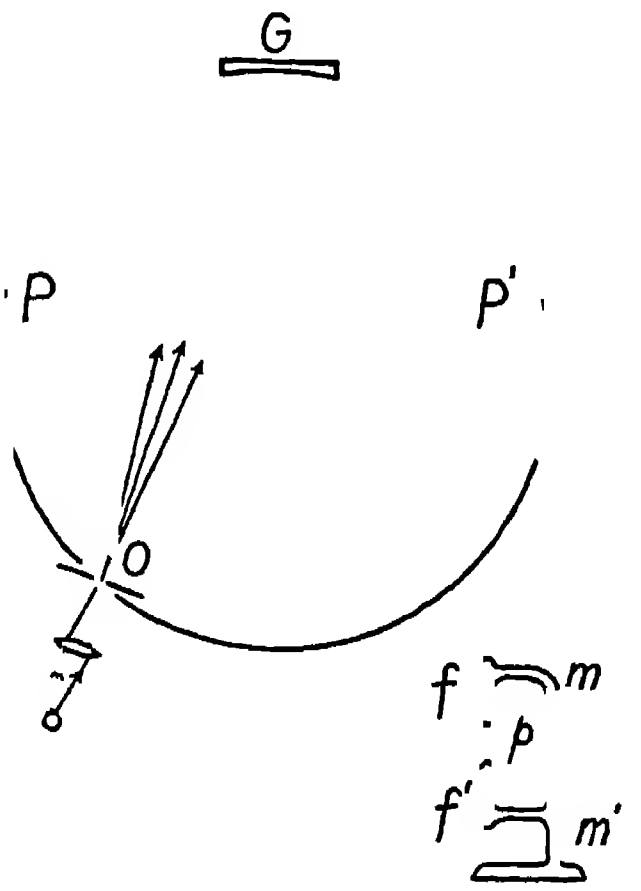


FIG. 137.—Paschen mounting

* Runge and Paschen, *Abhand. d. K. Akad. d. Wiss. z. Berlin*, Anhang, p. 1, 1902.

With the Paschen mounting the spectrum is in general not normal. It is normal only in the region directly opposite the grating, and it is only here too that the diffracted beam is incident perpendicularly upon the focal curve. For example, at P or P' the beam falls quite obliquely upon this curve. This is somewhat of a disadvantage in two ways. First, in making observations with an eyepiece, only so to say one spectrum line at a time is in focus. Second, in photographing, if the surface of the plate is not flat but is wavy, there will be small but appreciable displacements of the spectrum lines from their normal positions. These displacements are, however, in general taken care of when a correction curve is plotted from the measurements made upon the plate.

The Paschen mounting demands a considerable amount of space. For a large grating an entire large room, with a solid steady floor, is required. To insure steadiness of the mounting, a circular cement pier is usually built for supporting the frame PP'. An additional small pier suffices as a base for the grating. The spectrograph is not easily housed, and for this reason the source of light is commonly housed instead, the spectrograph proper being left open. This procedure in turn requires that provision be made for darkening the room completely. Moreover, the installation must be made where there are no marked temperature fluctuations. Such fluctuations cause spectrum lines to shift position during long exposures leading to a broadening of the photographic images. Often too they cause a shift in the relative positions of the comparison spectrum and the unknown spectrum, even when, as is common in practice, the comparison spectrum is thrown on to the plate both before and after the exposure of the unknown. Change of temperature is deleterious mainly because it causes expansion or contraction of the grating and thus causes the grating space to alter. Change of temperature of, for example, the circular frame, or of the photographic plate, is relatively innocuous. Consequently much can be done to lessen the harmful effects of temperature change in the room by housing the grating alone, leaving only the necessary aperture for passage of the incident and emergent beams. The housing must be lagged or controlled in temperature by a thermostat.

When the Paschen mounting is used, two or more slits may be placed at different points along the Rowland circle in order to facilitate working at different angles of incidence.

29. The Abney Mounting.*—Referring to Figure 138, the plate-holder or camera, C, is situated on the normal to the grating, G. The photographic plate, P, is bent to the curvature of the Rowland circle. The slit is erected at some position, O, on this circle, so chosen that the desired spectral order

* *Phil. Trans.*, Part II, p. 457, 1886.

and region fall into the camera. To pass from one spectral region to another the slit is moved along the circle as may be necessary. Some refocusing is required each time. The slit is usually mounted at the end of a radius arm, A, which has a length equal to the radius of the reference circle and is pivoted at the center of this circle.

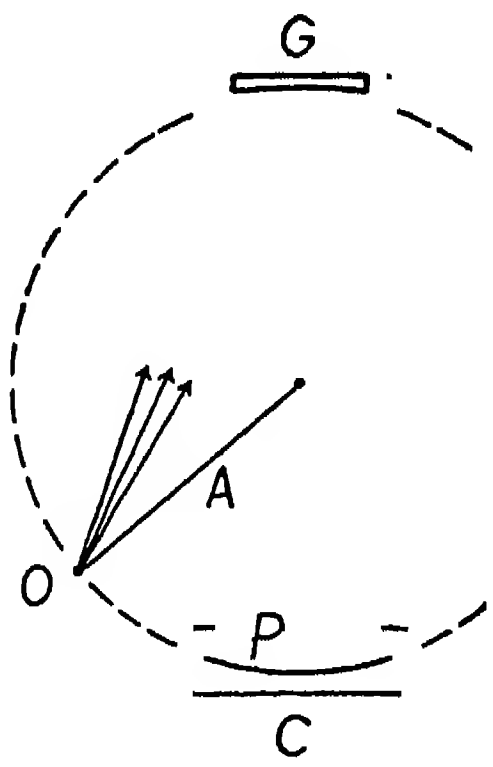


FIG. 138.—Abney mounting

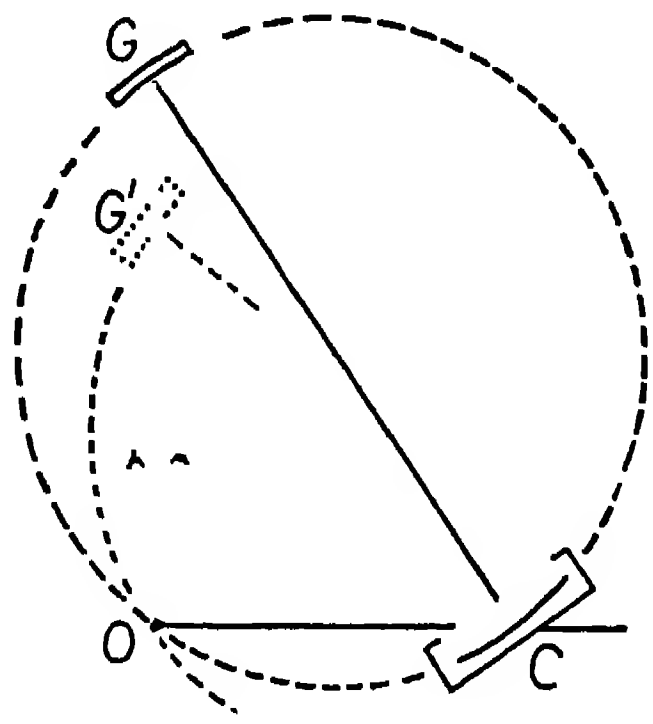


FIG. 139.—Rowland mounting

The spectrum is normal in the region where observation is carried out, and, moreover, in this region the beam falls perpendicularly upon the focal curve. These advantages led to a former popularity of the Abney mounting.

The demands made by the Abney mounting resemble those made by the Pachen mounting. However, the Abney mounting admits of housing fairly easily.

30. The Rowland Mounting.*—The mounting upon which Rowland decided for his own use is represented in Figure 139. It involves a motion of the grating, G, and of the camera, C, in changing from one spectral region to another. The slit is situated at O and remains fixed in position, and the incident beam remains fixed in direction. The grating and camera are mounted at opposite ends of a trussed girder, GC—at opposite ends of a diameter of the Rowland circle. The lines GO and OC are, together, inscribed in a semicircle, and hence these lines make a right angle with each other. Making use of this fact, a track is laid along GO and another along OC and each end of the girder GC rests upon a car, one car running on each track. The ends of the girder are pivoted over the centers of the cars. When it is desired to change from one spectral region to another, the end C, of the girder, carrying the camera, is moved, let us say away from O, whereupon the end G, carrying the grating moves, let us say, to G'. When the grating moves we

* *Phil. Mag.* (5), **16**, 197, 1883, or *Physical Papers*, No. 30, p. 492.

think of the Rowland circle as moving with it and, under the given mechanical conditions, the circle moves in such a manner that the slit lies always upon it. The circle slides, so to speak, through the slit, changing its inclination with the face of the slit in the course of the motion.

In the Rowland mounting the spectrum is normal in the region where it is observed, and in this region also the beam falls perpendicularly upon the focal curve. Further, as previously stated, the slit and the incident beam remain fixed in space. Rowland, using this mounting, carried out extensive investigations upon the solar spectrum, and in these it was a great advantage to have the slit and incident beam remain fixed. Rowland adopted the mounting which now bears his name because of a desire to combine the various aforementioned advantages. The mounting, however, involves a considerable amount of mechanism and when spectrographs are contemplated which are to be used only for the investigation of the spectra of portable terrestrial sources of light, this mounting is now not often chosen. In addition to the difficulty of mechanical construction, a drawback of the Rowland mounting lies in the fact that the motions of the grating and camera introduce difficulty in maintaining adjustments which are not encountered in the Paschen and the Abney mountings.

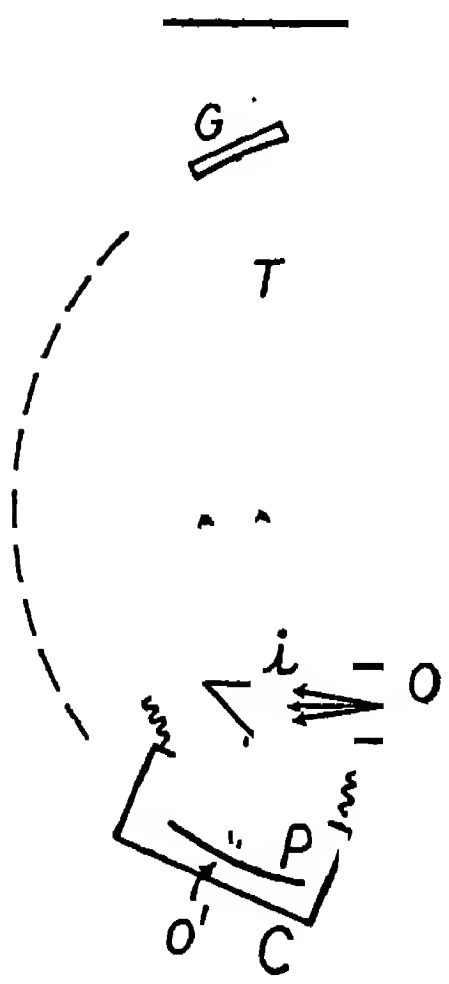


FIG. 140.—The Eagle mounting.

The Rowland mounting and the Abney mountings resemble each other closely. They are alike except for the fact that what moves in the one remains fixed in the other, and vice versa.

31. The Eagle Mounting.*—This mounting is represented in Figure 140. The slit is located at O and a small internally reflecting prism, i, directs the light upon the grating, G. The virtual image of the slit formed by reflection in the prism falls at O'. The camera, C, holds the photographic plate, P. The slit and reflecting prism are situated just below the central horizontal plane through the grating, and the photographic plate is situated just above this plane, and consequently the prism does not obstruct any part of the beam returning from the grating to the photographic plate.

To pass from one spectral region to another, the grating, G, is moved along the track, T, and also rotated by the proper amount to cause the Rowland circle, which we must think of as moving with the grating, to pass always through O'. The camera is pivoted about an axis through O' and must

* *Astrophys. Jour.*, **31**, 120, 1910.

be given a corresponding rotation in order that the surface of the plate may always lie along the Rowland circle.

The Eagle mounting is an adaptation of the Littrow principle to the concave grating. The distinguishing feature of the Littrow mounting of a prism or plane grating is that the return beam doubles back on the incident beam, and this is also the distinguishing feature of Eagle's concave-grating mounting.

The slit, instead of being placed at O, may alternatively be placed at O', the small prism being dispensed with. The slit in this case again lies slightly below the central horizontal plane and the photographic plate slightly above this plane.

The Eagle mounting requires much less space than the other mountings which have been described and is much more easily housed, and when housed the temperature may be more easily controlled. Moreover, with this mounting the astigmatism is less than half as great, at a given wave-length and spectral order, than in the Abney or Rowland mountings. (This fact is demonstrated in Appendix E.) Furthermore, higher spectral orders can be attained than with the Abney or Rowland mountings.*

The one drawback of the Eagle mounting is that it involves both translation and rotation of the grating and in addition a rotation of the camera. It is apparently this fact which has prevented the mounting from receiving as wide an adoption as it otherwise so well merits; but it is being increasingly used. The spectrograph is ordinarily provided with a linear scale along the track and with circular scales at the grating and at the camera. Corresponding settings of these three scales for each of numerous contiguous or overlapping spectral regions are determined and recorded in advance of undertaking investigation. The record then serves as a basis for future setting in any desired spectral region. Some refocusing by trial is necessary, however, after each change—except when the mechanical parts are extremely well made.

32. The Mounting in Parallel Light.†—This mounting which, for reasons which we shall learn, is also called the *stigmatic mounting*, is represented in Figure 141. It is the only well-known mounting which is not based on the principle of the Rowland reference circle. Light from the slit, which is located at O, is reflected by a concave mirror, M, and, leaving this as a parallel beam, is incident upon the grating, G. The camera, C, holding the plate, P, is situated on the normal to the grating. The grating is mounted

* No corresponding simple comparisons with the Paschen mounting are possible.

† F. L. O. Wadsworth, *Astrophys. Jour.*, **3**, 47, 1896; in particular pp. 55 ff. See also Fabry and Buisson, *Jour. de. phys.*, **9**, 929, 1910; Meggers and Burns, *Bull. Bur. Stand.*, **18**, 185, 1922, where details of construction and further references are given.

at one end of an arm, GB. This arm is pivoted about a vertical axis at the end G, the pivot being located under the grating. The arm is thus free to rotate about the end G, in the plane of the figure, and when the arm rotates, the grating rotates with it. To pass from one spectral region to another the end B of the arm is caused to slide over the ways WW'. The camera rests on the arm and is carried with it. With each change of position of the arm the distance from the grating to the photographic plate must be altered in order to maintain the focus. Accordingly, provision is made for sliding the camera lengthwise along the arm GB.

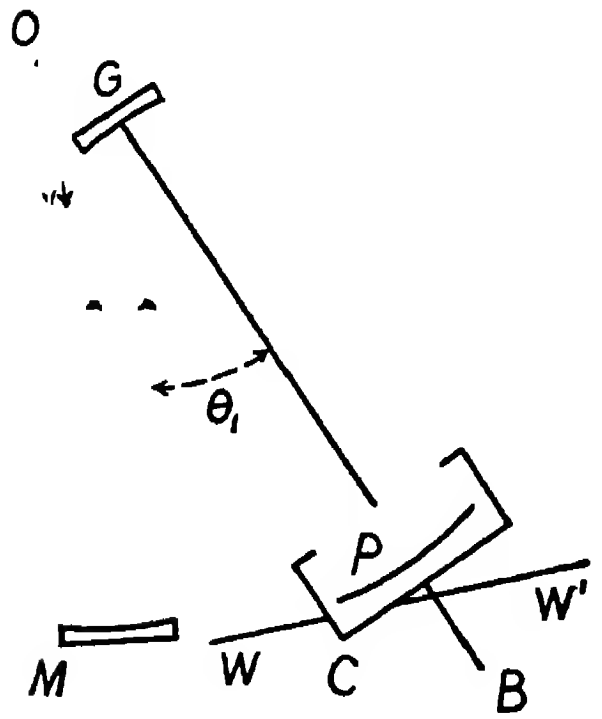


FIG. 141.—The mounting in parallel light.

The distance from the grating to the point along its normal where the diffracted rays are focused is a function of the angle of incidence upon the grating, θ_1. This distance is derivable from equation (19). Divide both numerator and denominator of this equation by r_1. Since the incident light is parallel, we must consider the virtual image of the slit formed by the concave mirror to lie at infinity. Accordingly, place $r_1=\infty$. Since the field point lies on the normal to the grating, the angle of diffraction, θ_2, is zero, or $\cos\theta_2=1$. We thus obtain for the distance from the grating to the photographic plate:

$$r_2=\frac{R}{\cos\theta_1+1}. \tag{20}$$

When the end B of the arm GB is moved toward W', the angle of incidence increases, and, in accordance with equation (20), the camera must be moved away from the grating.

In designing a mounting of the type in question, provision should be made for a variation of the angle of incidence in the range from about zero to 45°. Correspondingly, r_2 will vary from a value of about $R/2$ to $.6R$. By way of comparison, mountings based upon the Rowland circle require provision for a maximum distance from grating to photographic plate of $r_2=R$. Thus the parallel light mounting is only a little more than half as large, in linear dimension, for a grating of a given radius of curvature, as are the mountings based upon the Rowland circle. The dispersion is reduced in proportion to the reduction of the distance from the grating to the photographic plate. The visual resolving power is unaffected, however.

In the present mounting the astigmatism is equal to zero on the normal to the grating. This is proved in Appendix E. Practically speaking, the astig-

matism is zero over the entire length of the photographic plate, and it is for this reason that the mounting is called *stigmatic*. The lack of astigmatism carries with it a gain in intensity in comparison with astigmatic mountings and there is a further gain in intensity because the grating, being used at a shorter image-distance, subtends a larger angular aperture than in other mountings. The over-all gain in intensity has been estimated as being between five- and tenfold. The lack of astigmatism furthermore makes it possible to use a comparison prism at the slit in the usual manner or to use complementary diaphragms at this point. The spectrum is normal.

The parallel light mounting was first conceived by Wadsworth, who, however, used a lens for collimating instead of a concave mirror. The mounting is not usually referred to as the "Wadsworth mounting" for fear of confusion with the well-known Wadsworth prism mounting. It has been used effectively by Runge and Paschen, by Fabry and Buisson, and at the United States Bureau of Standards by Meggers, Burns, Kiess, and others (for references see previous footnote).

33. Mountings for the Extreme Ultra-violet.*—The extreme ultra-violet region of the spectrum is defined, somewhat arbitrarily, as extending from about 1850 Å to about 140 Å. This is a region of great opacity for all known solids—excepting only fluorite which is transparent down to 1200 Å. Even most gases are highly opaque, and hence a well-evacuated spectrograph is required.

The first penetration into the extreme ultra-violet was made by means of a prism spectrograph—a diminutive vacuum spectrograph provided with a fluorite prism and fluorite lenses. The instrument in question and the special photographic plates which are necessary for work in the extreme ultra-violet were developed by V. Schumann, whose labors bore fruit principally during the decade 1890–1900. Schumann explored the spectral region 1850–1200 Å, which is now known as the "Schumann region." However, since he used a prism spectrograph, he could not measure the wave-lengths of the spectrum lines which he photographed. But he estimated the wave-lengths of the lines which he found, by extrapolating the dispersion curve for fluorite from the ordinary ultra-violet into the Schumann region.

From about the year 1906 onward, T. N. Lyman applied the concave grating to the Schumann region and was able to determine wave-lengths with much-needed improvement in accuracy. Lyman, moreover, extended the known ultra-violet from 1200 Å to about 500 Å.

* For further information see Theodore Lyman, *The Spectroscopy of the Extreme Ultra-violet* (Longmans, Green & Co., 1928), where a Bibliography is given through 1927. See also Thomas H. Osgood on the related subject, "The Spectroscopy of Soft X-rays," *Phys. Rev. Suppl.*, **1**, 228, 1929.

The concave-grating mountings which are used today in the extreme ultra-violet are of two principal types. The older type, represented in Figure 142, is the one which was used by Lyman. A heavy brass tube, T, having a diameter of some 15 or 20 cm, serves as a housing. This tube is provided with flanges to which end plates, E, are sealed and bolted, and the spectrograph is highly evacuated. The same type of mounting was also used by Millikan and Sawyer and Bowen in their marked further extension of the spectrum to 140 Å. The distinguishing feature of this type of mounting is that the radiation falls upon the grating at a small angle of incidence and is photographed at small angles of diffraction.

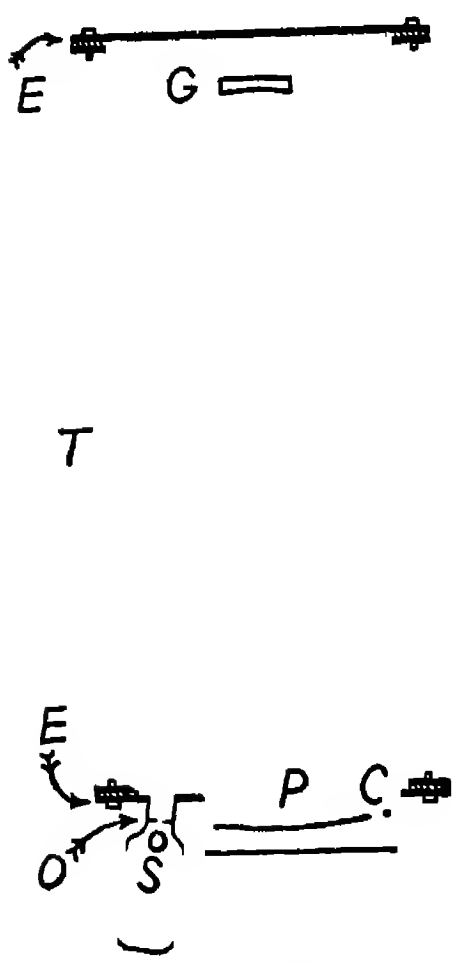


FIG. 142

In the figure, S represents the source, which consists either of an electric discharge in a gas at low pressure, or of a short spark in high vacuum, a so-called "hot spark." The source is placed close to the slit, O, in order that the light passing through the slit may fill the grating, G, as nearly as possible, and in order that as little radiation as possible may be absorbed in the small chamber containing the source, before entering the slit. The state of evacuation of the spectrograph is controlled by two pumps, one connected to the chamber containing the source and the other to the main chamber. The two chambers connect only through the slit, O, which is very narrow, perhaps .01 mm wide, and often very short, perhaps only .2 mm long. The source chamber usually contains much more residual gas than does the main chamber.

The central image formed by the grating falls, let us say, at C, to the right of the photographic plate, P, which is some 8 or 10 cm long. The plate receives that part of the first-order spectrum extending from let us say 150 Å, falling at the right-hand end of the plate, to 1500 Å falling at the left end. The major portion of the extreme ultra-violet region can be covered at a single setting of the spectrograph. The remainder can be reached if desired by turning the grating slightly and then resetting the plate-holder to the position of best focus.

The gratings used as a rule have about 15,000 lines per inch, although 20,000 lines and 30,000 lines per inch have been used. The ruling must be very light. Moreover, glass has a higher reflecting power than speculum in the extreme ultra-violet and is less subject to deterioration caused by the products of the electric discharge; accordingly gratings ruled upon glass are frequently used. They are originally ruled with an even more shallow groove than is finally desired, as this procedure avoids the appreciable wear of the

diamond which otherwise occurs when ruling on glass. After ruling, the grooves are etched deeper in dilute hydrofluoric acid by a process developed by R. W. Wood.* The gratings used in the extreme ultra-violet are not of the largest—a ruled surface of 8 cm wide by 5 cm high is about the maximum. A radius of curvature of 1 m has been the most common in the type of mounting which we have been discussing—however, gratings having radii of as high as 7 m are now coming into use.

The photographic plates, the so-called "Schumann plates," differ from ordinary plates principally in having a minimum amount of gelatin present in the sensitive coating. Gelatin absorbs the extreme ultra-violet and when present in too great quantity prevents the radiation from reaching the grains of silver bromide.

Various individual spectrographs, even when they are of the same general type, may differ markedly from each other in construction. The differences have to do mainly with the manner of mounting and housing the source and photographic plate. Such details assume importance when, as in the extreme ultra-violet, the technique of investigation is elaborate and difficult.

In the newer type of mounting (see Fig. 143) the radiation falls upon the grating at a large angle of incidence. This feature was introduced in the year 1927 in so far as the *concave* grating is concerned. The method of grazing incidence had, however, been previously developed by A. H. Compton and others for the diffraction of a narrow pencil of *X-rays* by a *plane* grating, and in the year mentioned an investigation was carried out by T. H. Osgood,† at Compton's suggestion, which showed that the grazing method also promised valuable results in the extreme ultra-violet when a concave grating is used.

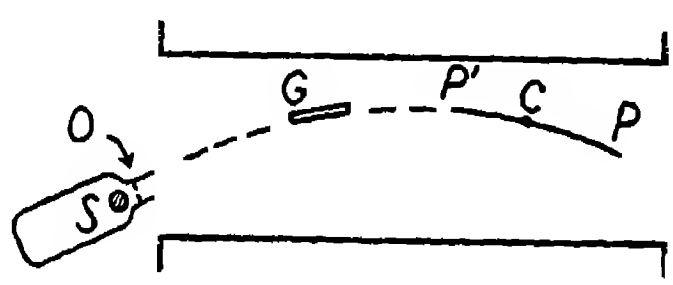


FIG. 143

The angle of incidence must be 80° or more, or, in other words, the angle between the beam and the surface of the grating, the so-called "glancing angle," must be 10° or less. The slit, O, lies on the Rowland circle near the grating, G, and the photographic plate, $P'P$, lies on the other side of the grating, also near it and on the Rowland circle. The central image falls at C and the spectrum is photographed on both sides of this, principally in the region CP but also along CP'. The dispersion is very great, especially along CP'. The extent of the spectrum on this side is definitely limited, however. The spectrum is far from being normal anywhere along the plate.

The grating used by Osgood had a radius of curvature of about 2 m. The

* *Phil. Mag.* (6), **48**, 497, 1924 (in particular p. 507), and also Wood and Lyman, *ibid.* (7), **2**, 310, 1926.

† *Phys. Rev.*, **30**, 567, 1927.

distances from slit to grating and from grating to central image were each 35 cm or less.

The advantage of the large-angle mounting lies in the large dispersion which is obtained at large angles and in the fact that the reflecting power of the grating (which in the extreme ultra-violet is less than 10 per cent at normal incidence)* is much greater at nearly grazing incidence. The large angle of incidence entails increased astigmatism of the grating and this, if it were the only factor which is operative, would result in a loss of intensity. The gain in reflecting power is, however, more than enough to offset this loss, leaving a considerable net gain in intensity at nearly grazing incidence. A phenomenon of total reflection occurs in the extreme ultra-violet when a beam is reflected *externally* at nearly grazing incidence which is analogous to the phenomenon of total *internal* reflection of ordinary optics. The same phenomenon occurs in the reflection of X-rays, only with these the critical glancing angle is even smaller. This matter is discussed in the chapter on the diffraction of X-rays (chap. 8, sec. 22).

Osgood conducted his investigation as an extension of the soft X-ray region from the then existing limit of 27 Å toward longer wave-lengths, that is, toward the then short wave-length limit of the extreme ultra-violet, 136 Å. He succeeded in bridging the gap which had existed up to that time, and in fact overlapped into the extreme ultra-violet region. It should be mentioned, however, that A. Dauvillier,† just before Osgood, had success in bridging a part of the gap in question, but by a quite different method—by a method which does not concern us here.

When the concave grating is used at a large angle of incidence the aberration becomes large (see Appendix D). It has recently been shown by Mack, Stehn, and Edlén,‡ and by I. S. Bowen,§ that in the Osgood mounting the optimum width of the grating may be quite small, perhaps only 1 cm.

ERROR OF RULING

(Secs. 34–43)

A diffraction grating must be ruled with great precision if it is to fulfil its function, and hence it becomes a matter of importance to know the effects of errors of various types.

34. General.—An *ideal grating* is defined as one having grooves which are straight and parallel, are equally spaced, and all have identical form. Identity of form includes exact equality of depth.

* P. R. Gleason, *Proc. Nat. Acad.*, **15**, 551, 1929.

† *Jour de phys. et le rad.*, **8**, 1, 1927.

‡ *JOSA*, **22**, 245, 1932.

§ *Ibid.*, **23**, 313, 1933.

The foregoing definition does not touch upon such questions as dispersion or intensity which are questions of great practical importance. Accordingly, the most nearly ideal grating is not always the best.

The theory which we shall develop applies equally well to the plane and to the concave grating. However, for the sake of simplicity we shall always speak in terms of the plane grating, it being supposed that lenses are used for collimating and focusing, in the usual manner. It will be assumed that the lenses are perfect and that the surface of the grating is accurately plane.

Corresponding to the desiderata that the grooves be straight and parallel, equally spaced, and all have the same form, we shall designate three primary classes of error:

1. Curvature and lack of parallelism of the grooves.

2. Error of spacing. (This is by far the most important of the three classes.) Any groove which is out of its proper position contributes a disturbance at a given point of the focal plane which is out of its proper phase. Accordingly, error of spacing is alternatively designated as *phase error*.

3. Variation of groove form. When the grooves differ from one another in form they scatter different amounts of light in a given direction, or, in other words, contribute disturbances which differ from one another in amplitude. Therefore this class of error is alternatively designated as *amplitude error*.

35. Curvature and Lack of Parallelism of the Grooves.—Several cases which fall into this class are discussed by Lord Rayleigh.* We shall here discuss only two cases—general curvature and general lack of parallelism—and these only briefly.

Slight general curvature of the grooves introduces only a small amount of astigmatism which is practically harmless. Let us see how the astigmatism arises. Suppose a parallel beam originating from a point in the slit of the spectroscope is incident perpendicularly upon the grating. Referring to Figure 144, a strip b of the grating acting alone gives rise to a dot d as central image and to a dot e as first-order spectral image on the left, whereas the strip c acting alone gives rise to d as central image and f as spectral image. The entire grating therefore yields a short line, ef, as the spectral image of a point of the source.

The diffracted wave front, as it leaves the grating, instead of being plane is a cylinder, with elements horizontal, slightly converging, for a spectral image on the left of the normal.

The astigmatism is rendered evident by the fact that the familiar dust marks which appear when the slit is almost closed, focus farther in, under the supposed conditions, than do the spectrum lines.

* *Op. cit.*, **3**, 112, or *Encyc. Brit.* (9th ed.), Vol. **24**, "Wave Theory of Light."

Similar slight astigmatism is introduced by one type of "error of run," as we shall learn later.

Rowland estimated, as the result of a test, that the lines of his gratings deviate from straightness by less than 1/100,000 inch. A considerably greater curvature than this amount could however be tolerated.

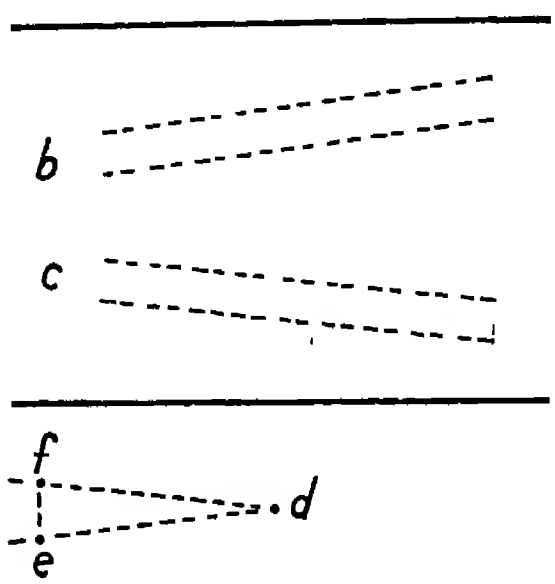


FIG. 144

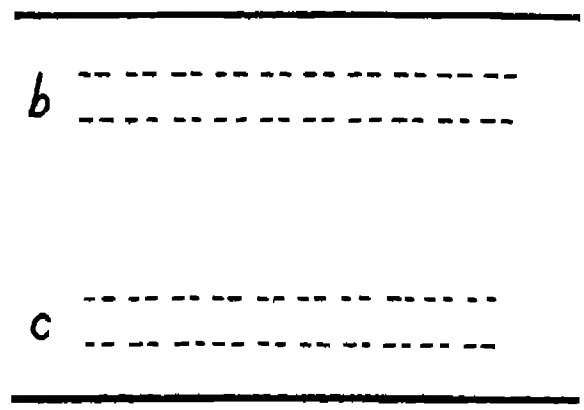


FIG. 145

Lack of parallelism of the grooves, as soon as it becomes appreciable, is a serious fault. Referring to Figure 145, the grating space which obtains for the strip *b* of the grating, acting alone, is greater than that for the strip *c*. Therefore the spectral image due to *b* lies nearer the central image than that due to *c*. Consequently, when the entire grating is effective at once, we have too broad a spectral image, entailing loss of definition and resolving power.

The grooves will be parallel only if the "ways" along which the grating carriage travels are perfectly straight. The ways may deviate from straightness by 1/100,000 inch without harmful result if this deviation arises from a uniform curvature. But a deviation from straightness of as much as 1/10,000 of an inch would be damaging when a large grating is ruled.*

36. Periodic Error of Spacing.—We shall now take up in succession various errors of Class 2, namely, "Spacing Errors," and among these shall consider first the most important of the *systematic* errors of spacing.

Periodic error of spacing is a systematic error and is important because it gives rise to false lines in the spectrum, so-called *ghosts*. These always exist; though with a good grating only intense spectrum lines show them, and then only faintly. This type of error does not affect the perfection of the principal spectral images or their positions.

Early gratings, those of Nobert and of Rutherfurd, had what we would

* Taking the length of the ways as 20 inches and assuming uniform curvature, which is the least harmful, a sagitta of .0001 inch leads to a radius of curvature of 5×10^5 inches, or about 8 miles (in place of an infinite radius). Simple considerations of geometry and grating theory now lead to the conclusion that the spectral images would be considerably broader than for a large ideal grating.

today call large periodic error, which arose perhaps mainly from error of the screw of the ruling machine. The slope of the thread was not strictly constant but varied slightly in a manner which was repeated with every turn. Rowland devised a method of cutting, and especially of finishing, screws so that they were practically free from error; but he had still to contend with periodic error introduced by imperfection of the bearings in which the screw is mounted, by imperfection of alignment of the axis of the screw thread with the bearings, by slight eccentricity of the spacing head, etc.

J. A. Anderson* has pointed out that *if* all parts of the machine were *truly rigid*, the periodic error could be held to such low limits that its effect could scarcely be detected, if at all. But a considerable force must be applied to overcome the friction of moving the grating carriage along its ways, and *elastic deformation* of the parts results. The over-all elastic deformation is 1/10,000 inch or more, or, for a grating of fifteen or twenty thousand lines per inch, is about $2a$, where a is the grating space. If, as the screw turns, the friction varies by as much as 1/200 of its value, the elastic deformation, varying correspondingly, will vary about 1/2,000,000 inch, or about $a/100$. This, as will develop, is sufficient to produce ghosts of intensity near the limit of what can be tolerated for a good grating. Thus we must always reckon with a periodic error which repeats itself with every turn of the screw or, what is the same, with every turn of the spacing head.

In Figure 146 the equally spaced tiny circles at the top of the figure represent the lines of an ideal grating, and the dots, in the next row, represent

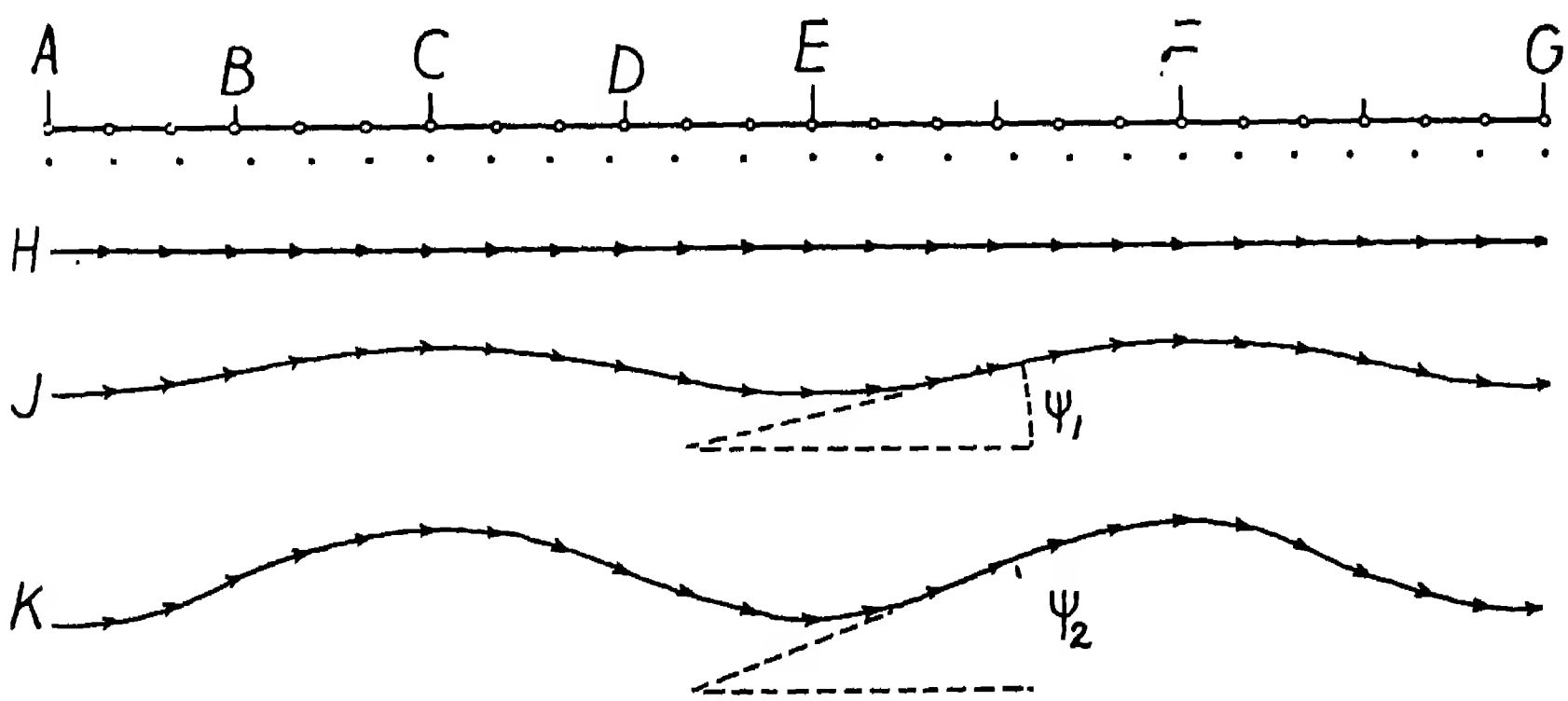


FIG. 146

the lines of an actual grating. At A the position is correct, at B it is too far to the right, at C it is correct, at D it is too far to the left, and at E it is again correct. We have now completed one period of the error—one turn of the

* *JOSA and RSI*, **6**, 434, 1922, and Glazebrook, *Dictionary of Applied Phys.*, **4**, 30.

spacing head. Another period, another turn, is represented from E to G. With a screw having twenty threads per inch, a grating five inches wide involves one hundred turns of the screw and spacing head.

In the same figure, H represents the vibration diagram for the ideal grating—at any spectral image, including the central image. This same diagram also obtains for the central image of the actual grating. At the central image the optical paths by way of the various lines of the grating are all equal; hence error of position of a groove does not alter the path and consequently does not affect the central image.

The curve J represents the vibration polygon, or curve, for the actual grating at the first principal maximum on the left of the central image. If the grating is to be a good one, the maximum slope of the curve, represented by the angle ψ_1, must be small. The maximum displacement of the grooves from their correct positions should not be greater than about $a/100$, and hence the angle ψ_1 should not be greater than about $.01 \times 2\pi = 3.6°$.

The curve K represents the vibration curve for the second order. A given displacement of a groove from its correct position now causes twice as great an error in phase angle. If the maximum displacement is $a/100$, the angle indicated by ψ_2, will be $.02 \times 2\pi = 7.2°$. The amplitude of curve K is twice as great as that of curve J. The effect of the periodic error increases with the spectral order.

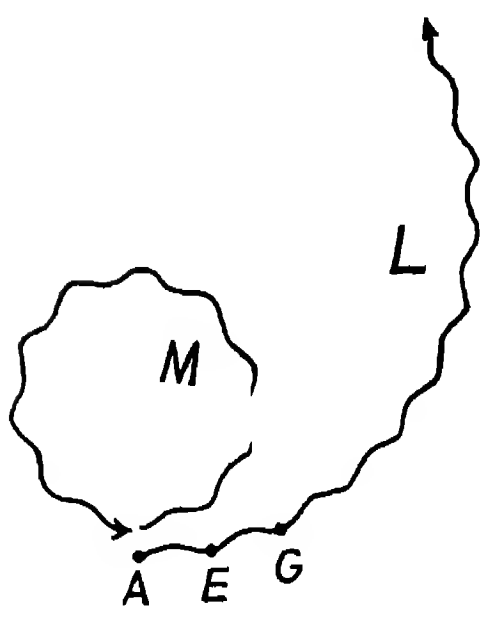


FIG. 147

Let us now consider what happens to the vibration curve as we leave the center of one of the principal maxima and pass, let us say, to the left. Referring to Figure 147, while we are on the slope of the maximum we have some such curve as L, in which the portions AE and EG each correspond to one turn of the screw. Proceeding further, the vibration curve will close when it makes one convolution, curve M, and it will likewise close when it makes two, three, etc., complete convolutions. The periodic error has practically no effect—until we reach a stage in which there is one convolution of the vibration curve for each portion of the grating ruled in one turn of the screw; then it does have an effect, as will develop in the following section.

37. Rowland Ghosts.—The stage in which there is one convolution of the vibration curve per turn of the screw is represented in Figure 148. The upper half of this figure applies on the left of the principal maximum, and the lower half of the figure applies on the right. In each case the curve ACE represents the first convolution of the vibration curve. On the left of the principal maximum, the curvature is less at C, upper right-hand corner

of the figure, than at A or E. As a result there is an error of closure AE represented by the vector s_1. On the right of the principal maximum the curvature is *greatest* at C, lower left-hand corner of the figure. The curves in the center of the figure represent the complete vibration curve in each case. What is of importance is that in each case the errors of

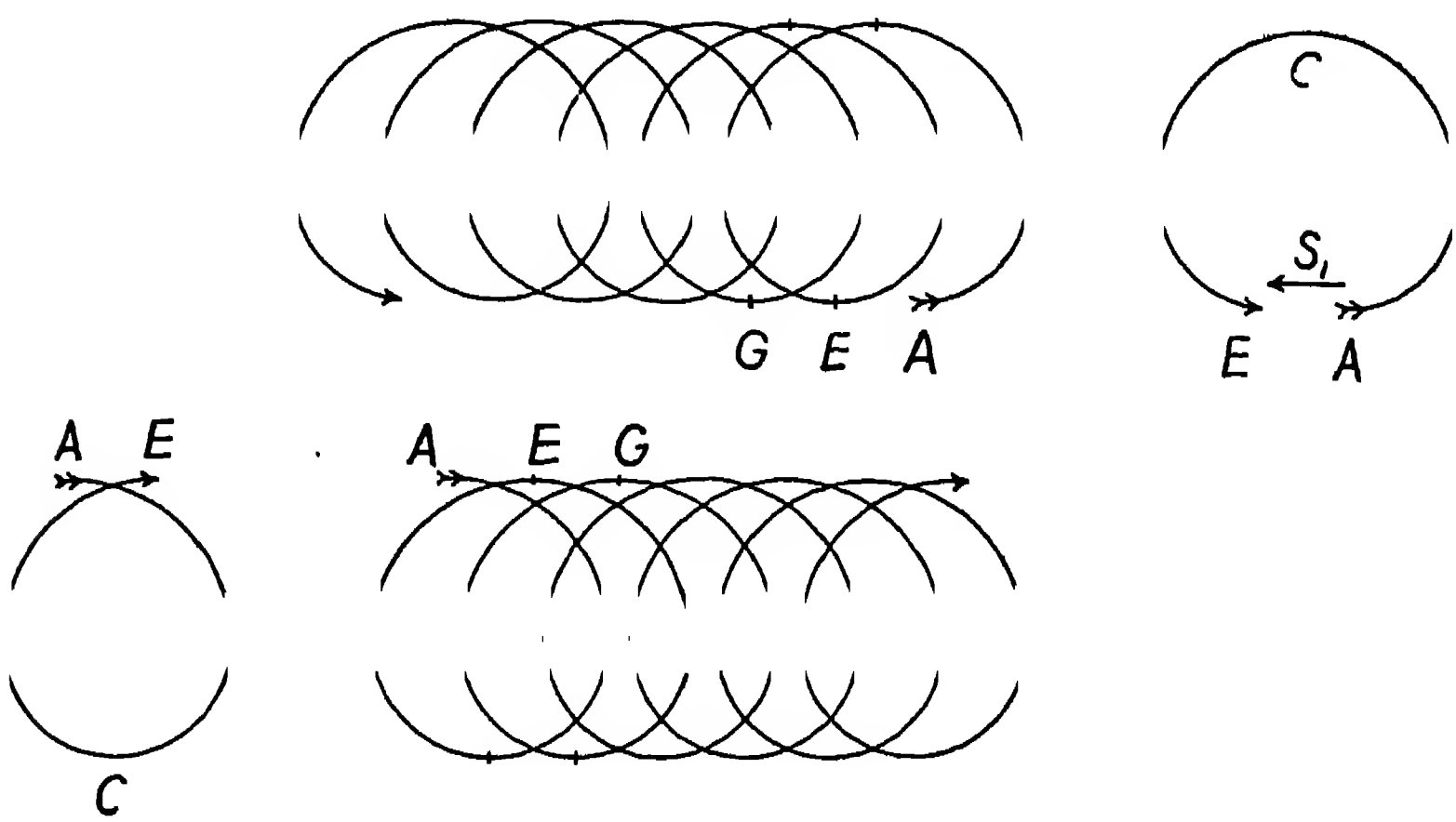


FIG. 148

closure, AE, EG, etc., of successive convolutions are all in the same phase and hence each set adds up to yield a false spectrum line, a so-called *ghost*. The ghosts are symmetrical with reference to the parent-line in the positions which they occupy and as a rule also in regard to their intensities, but not always. If n' be the number of turns of the screw involved in ruling the grating—supposing for convenience of discussion that this be integral—and s_1 be the error of closure per convolution of the vibration curve, then the intensity of the ghost will be $(n's_1)^2$. If the parent-line be in the first-order spectrum and the vibration curve make only one convolution per turn of the screw, then we will be dealing with the so-called ghost of the first order of a spectrum line of the first order.

In the second-order spectrum the ghosts of the first order are formed in the same general way as in the first-order spectrum, only now, when the vibration curve makes one convolution per turn of the screw, the error of closure per convolution is twice as great as before. This follows because the phase error for each individual groove is now twice as great. Thus the error of closure per convolution is $2s_1$ and the intensity of the ghosts is $(n' \times 2s_1)^2$, or four times as great as before. In the same way in the spectrum of order m, the intensity of the ghosts will be $(n'ms_1)^2$. That is, the intensity of the ghosts increases proportionately to m^2, the square of the order of the

spectrum. This refers to the intensity of the ghosts relative to the intensity of the parent-line, the latter in general changing from one spectral order to another.

The distance of the ghosts from the parent-line is the same in the spectra of different orders. Each ghost lies where we should have the minimum of number n', where n' is again the number of turns of the screw involved in ruling the grating. If n is the total number of lines of the grating, we have $n/n' = N$, where N is the number of lines ruled per turn of the screw, for example, 1000. When the vibration polygon makes one convolution per turn of the screw, the phase change is 2π per turn of the screw or *per N lines* of the grating. Now, in passing from one principal maximum to the next the phase change is 2π *per line*. Hence the distance of these ghosts from the parent-line is the $1/N$ part of the distance between principal maxima, for example, the 1/1000 part.

At the distance $2/N$ from the parent-line, in the spectrum of each order, the vibration curve makes two convolutions per turn of the screw and we must in general suppose that these two convolutions may exhibit an error of closure. Hence there may be ghosts of the second order at the distance $2/N$ on each side of the parent-line.

In the same way, at the distance $3/N$, when the vibration curve makes three convolutions per turn of the screw, there may be ghosts of the third order. Similarly, at the distances $4/N$, $5/N$, etc., there may be ghosts of the fourth order, fifth order, etc.*

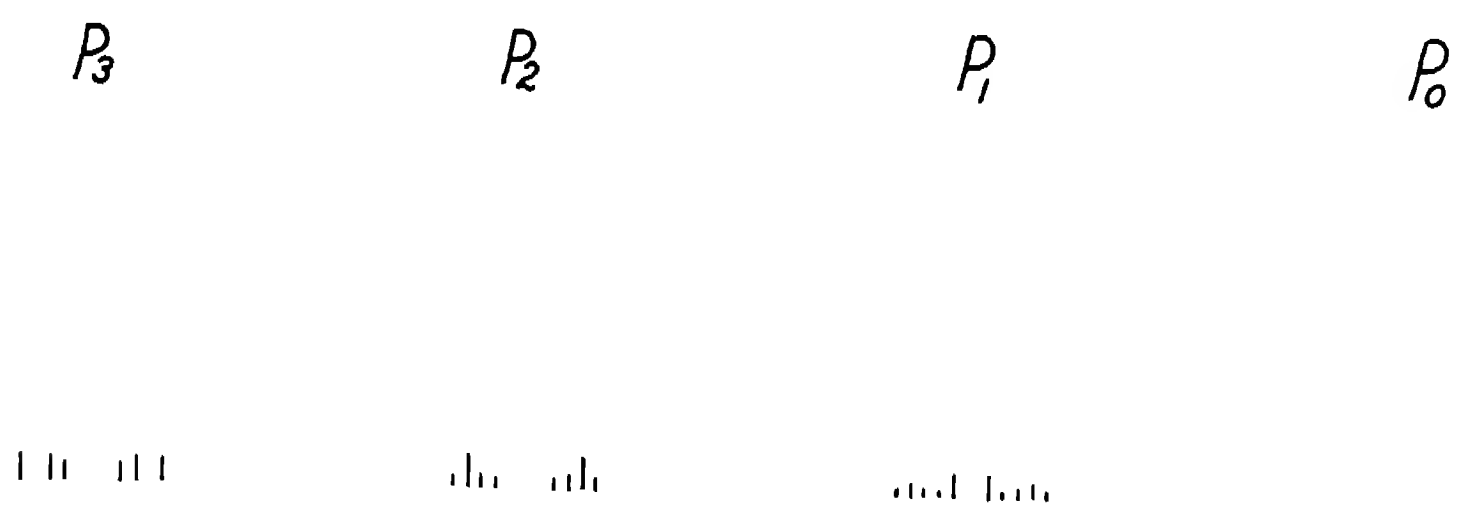


FIG. 149

Figure 149 represents a typical case of various ghosts of one parent-line in different spectral orders. The central image is P_0; there are no ghosts present there. The first-order image is P_1, the second-order is P_2, etc. The

* We need consider only integral numbers of convolutions of the vibration curve corresponding to one turn of the screw or fundamental period of the error. If the number is non-integral, the successive closing vectors will have progressive differences of phase and will fail to yield an appreciable sum, in the same way that the disturbances from the grooves of an ideal grating fail to yield an appreciable sum except when the phase differences are integral multiples of 2π, i.e., at the principal maxima.

intensities of the ghosts relative to the parent-line and their distances from the parent-line are exaggerated for purposes of representation. Ghosts of marked intensity are to be expected only in high spectral orders. For a given grating the ghosts of different orders have always the same relative intensity among themselves in the spectra of different orders. However, for different gratings there is no constant relation between the intensities of the ghosts of various orders. The first-order ghosts are usually the most intense. The next in intensity are often those of the third or the fourth order, most often the latter, as illustrated.

Ghosts of the type which we are considering were discovered in 1872 by G. Quincke,* who had at his disposal principally gratings ruled by Nobert. The theory of their origin was first treated in 1879 by C. S. Peirce,† who referred to the ghosts shown by Rutherfurd's gratings. Rowland, in building and perfecting his ruling machines, materially reduced periodic error and thus reduced the intensity with which the ghosts appear. He also greatly elaborated the theory.‡ Michelson later contributed a further elaboration.§

Ghosts of the type under consideration have today become known as "Rowland ghosts." They are characterized by the fact that they lie close to the parent-line and are symmetrical with respect to it. These features make it easy to identify them. They are further characterized by the fact that they originate from a periodic error having a fundamental period equal to one turn of the screw.

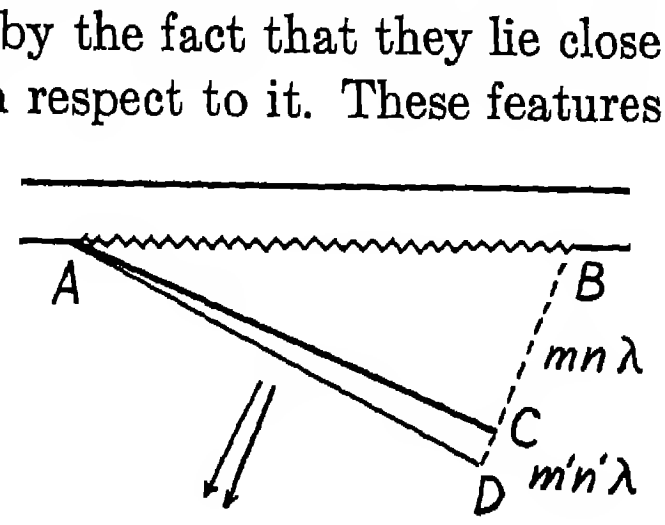


FIG. 150

Let us now derive an equation for the apparent wave-length of any Rowland ghost. Reference will be to Figure 150, which is drawn on the basis of normal incidence; the result will apply, however, for any direction of incidence. The symbols will have the following meanings:

λ wave-length of parent-line
λ' apparent wave-length of ghost
m, order of spectrum
m', order of ghost
n, number of lines of grating
n', number of turns of screw
N, number of lines per turn; hence $n=n'N$

The diffracted wave front for the parent-line is AC; thus $BC=mn\lambda$. The diffracted wave front for the ghost is AD; thus $CD=m'n'\lambda$. Hence $BD=$

* *Pogg. Ann.*, **146**, 1, 1872; in particular p. 57.

† *Amer. Jour. Math.*, **2**, 330, 1879.

‡ *Physical Papers*, p. 525.

§ *Astrophys. Jour.*, **18**, 278, 1903.

$mn\lambda + m'n'\lambda$. The apparent wave-length of the ghost is that wave-length, λ', for which the wave front AD is the same as for the ghost. The value of λ' is determined by the fact that:

$$BD = mn\lambda' .$$

Hence $mn\lambda' = mn\lambda + m'n'\lambda$, or, introducing an alternative minus sign to include also ghosts formed on the other side of the parent-line:

$$\lambda' = \lambda \left(1 \pm \frac{m'n'}{mn}\right) . \tag{21}$$

Or, since $n/n' = N$, we have:

$$\lambda' = \lambda \left(1 \pm \frac{m'}{mN}\right) . \tag{22}$$

If for a given grating N is not known, it can be determined.*

The ghosts may be regarded, in so far as their positions are concerned, as being formed by superposing a grating having one groove per turn of the screw upon an ideal grating having the same grating space as the actual grating. Each parent-line is then to be regarded as a central image for the coarse superposed grating. The ghosts occupy the positions which they would occupy if the actual grating had one extra line ruled at each complete turn of the screw.

When the eye is placed at one of the spectral images of, for example, the green mercury line, the periodic error of the grating is strikingly revealed to observation.

The relative intensities of the ghosts of different order are determined by the nature, the form, of the periodic error. The connection involved can best be made clear by taking a new point of view regarding the effect of the error of position of each groove. Referring to Figure 151, suppose the vector A' represents in amplitude and phase the disturbance contributed at a given point of the focal plane by a given groove of the actual grating. The vector A may be taken to represent the disturbance which would be contributed by this groove if it were in its correct position. Since the error of position of all grooves is small, the effect is to add small vectors such as v perpendicular-

* The number of lines ruled per turn of the screw is equal to the grating space divided into the pitch of the screw. The grating space is known or can be determined from a known spectrum line. The pitch of the screw will in general be one-twentieth of an inch, but if this be in question, it can be determined from the grating itself by a method described by R. W. Wood (*Phil. Mag.*, **48**, 497, 1924). This method is a photographic one. An alternative, visual adaptation is to place a scale some distance in front of the grating and focus on this with a small telescope placed in the beam.

ly along the vibration polygon for an ideal grating. Thus when the vibration polygon makes one convolution per turn of the screw, we may represent each convolution as shown in Figure 152. The closed equilateral and equiangular polygon ACE is for the ideal grating. The small vectors erected perpendicularly along this represent the effect of the errors of position of the rulings. Adding up the small vectors we obtain,

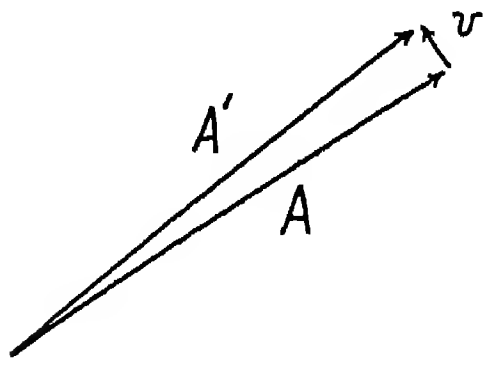


FIG. 151

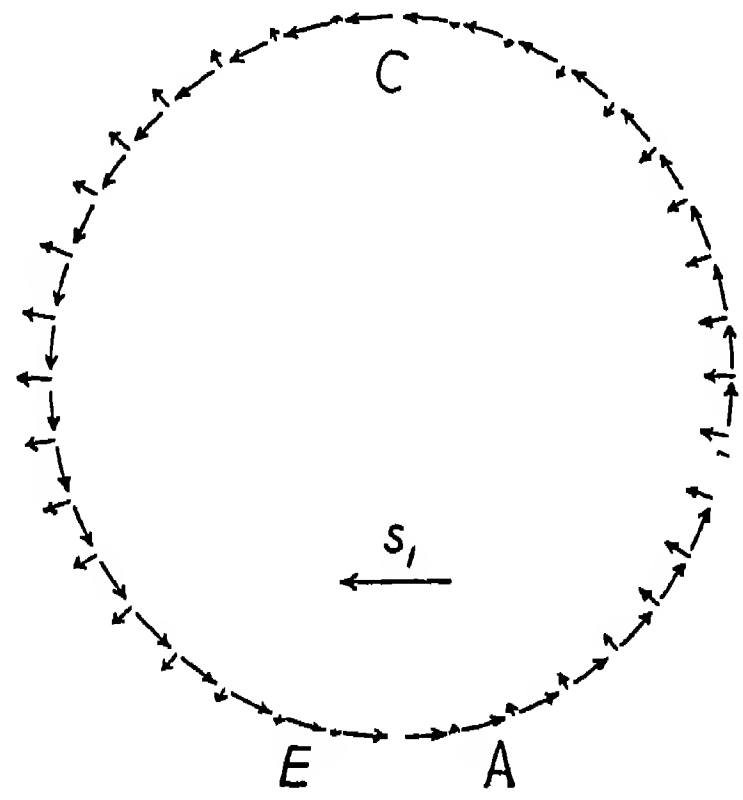


FIG. 152

overlooking inconsistencies of drawing, a vector s_1 which is identical with the vector s_1 of Figure 148 representing the error of closure of one convolution of the vibration curve, or polygon.

When the vibration polygon makes two convolutions per turn of the screw we have Figure 153. The perpendicular vectors point inward during the first convolution and outward during the second one. If the periodic error is a pure simple harmonic one, the sum of the small vectors will now be zero. In other words, the ghosts of the second order will vanish. Furthermore, it is obvious that for four, six, etc., convolutions of the polygon per turn of the screw, the sum of the small vectors will be zero. For three, five, etc., convolutions per turn of the screw the sum will also be zero, but this requires proof. (The proof is given in Appendix F.) Accepting the fact, we may say by way of conclusion that if the error is a pure simple harmonic one, with a period equal to one turn of the screw, we shall have ghosts of the first order, and of this order only—the ghosts of higher order will all vanish.

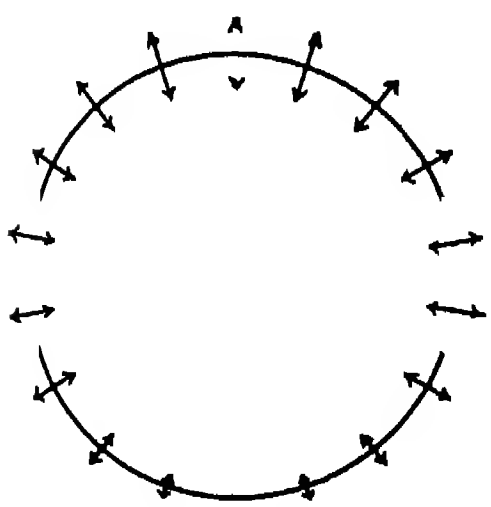


FIG. 153

It follows that when the error is not simple harmonic, then the ghosts as actually revealed by the grating represent by their intensities a Fourier analysis of the periodic error. The intensity of the first-order ghosts is a measure of the amplitude of the fundamental period; the intensity of the second-order ghosts is a measure of the amplitude of the first harmonic, etc.

The foregoing conclusions are more fully brought out in Appendix F, where, in addition, an equation giving the intensity of the ghosts in terms of the amplitude of the error is derived, and the magnitude of the permitted error is discussed.

A careful study of the ghosts formed by several gratings has recently been made by W. Rudolph.*

The extra-focal images formed by gratings are of interest in connection with the subject of error of ruling, especially in connection with periodic error. These images have been studied by R. W. Wood† and by S. Fagerberg.‡

38. Lyman Ghosts.—In the year 1901 T. N. Lyman§ discovered ghosts of a type which are characterized by the fact that they lie far from the parent-line and originate from a periodic error which is not associated with the period of the screw. These ghosts are so faint that they ordinarily cause no trouble, but under certain spectrographic conditions to be discussed later they are likely to be mistaken for real spectrum lines, and they hence demand attention.

We shall consider a hypothetical case chosen with a view to the possibility of drawing corresponding figures. The ghosts may be represented as shown in Figure 154. Here P_1 and P_2 are respectively the principal maxima

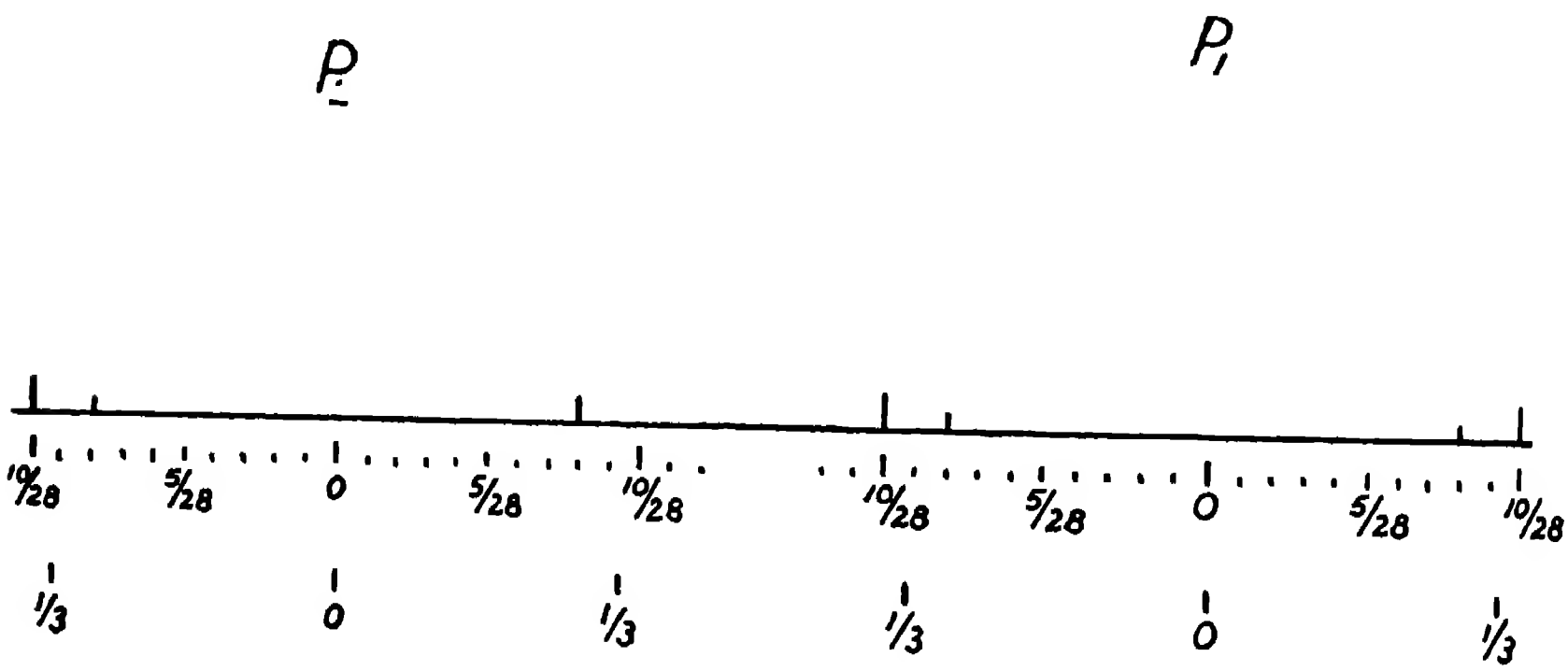


FIG. 154

of the first and second order of a parent-line λ. The ghosts are shown in groups of three to the right and left of each principal maximum, each group centering roughly, but not exactly, about the points which are distant from

* *Zeit. f. Physik*, **82**, 372, 1933.

† *Phil. Mag.* (6), **48**, 497, 1924.

‡ *Ibid.* (7), **5**, 204, 1928.

§ *Phys. Rev.*, **12**, 1, 1901, and **16**, 257, 1903.

the principal maxima by one-third the distance between principal maxima, as indicated.

Applying what we know about the formation of Rowland ghosts to the present case, we reason as follows: The ghosts in question may originate from some periodic error, of as yet unspecified origin, the error having such a nature, such a form, that the fundamental and the harmonics of low number are all weak, so weak that the ghosts of corresponding order are too faint to observe. But, for some reason, several of the harmonics of high order, and presumably of consecutive order, are relatively strong—strong enough to produce the observed ghosts. Now we know that ghosts of consecutive order are separated by $1/N$ of the distance between principal maxima, where N is the number of lines ruled per fundamental period of the error. In the example given the distance between the ghosts is the 1/28 part of the distance between principal maxima, hence on the foregoing basis the fundamental period of the error would be $N=28$. Moreover, the distances of the ghosts from the parent-lines are respectively 8/28, 9/28, and 10/28, of the distance from P_1 to P_2, hence the strong harmonics are those having frequencies respectively 8, 9, and 10 times that of the fundamental.

The fundamental period of the error, in our case $N=28$, and the order numbers of the strong harmonics, 8, 9, and 10, are all determined in a somewhat involved way by the driving mechanism of the ruling machine. Referring to Figure 155, suppose M is the motor pulley, B is the belt, and D is the driven pulley on the driving shaft (in Fig. 117, representing the ruling machine, this shaft is numbered 13). The belt is made as uniform in thickness and flexibility as possible, but it cannot be made perfectly uniform. Any outstanding non-uniformity will introduce periodic variation in the strain on the machine, and periodic error will ensue. The pulley D makes in our case about three revolutions for one revolution of the belt, but not exactly three; if it made exactly three revolutions the remaining non-uniformity of the belt would introduce a periodic error having a period $N=3$ grooves and there would then be relatively strong ghosts at the points exactly 1/3 of the distance between principal maxima.* In order to avoid these relatively strong ghosts, the belt is made of such a length that it makes one revolution for a number of revolutions of the pulley D differing from an integral number. In our case the number would be 3 1/9; in an actual case the difference from an integral number would be much less than 1/9. But in any case, if

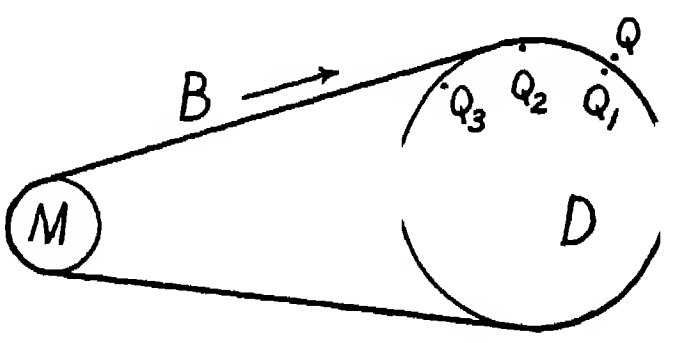


FIG. 155

* See *Rowland's Physical Papers*, p. 536.

Q and Q_1 are points respectively on the belt and pulley which are in contact at some arbitrary zero stage, then after one revolution of the belt the point Q will coincide with some point Q_2, and after two revolutions with some point Q_3, etc., until Q again coincides with the original point Q_1 of the pulley. Let us suppose that exact coincidence of Q and Q_1 recurs after nine revolutions of the belt or, in other words, after $9 \times 3\ 1/9 = 28$ revolutions of the pulley. There is one line of the grating ruled for every revolution of the driving shaft carrying the pulley, hence there would be 28 lines ruled. One fundamental period of the error would now be completed, $N = 28$. (In an actual case the number of lines ruled before the original coincidence recurs is several times larger. However, sooner or later the original coincidence must recur, practically speaking. The highest recorded case is $N = 348$.) To continue, the strongest Fourier component of the error will, in our case, be that multiple of 1/28 which has the value nearest 1/3, this being 9/28, and the corresponding ghost will be the strongest, as illustrated. This will appear more fully in the following paragraphs. Next in order of strength will be 10/28 and 8/28.

Referring to Figure 156, suppose the row H of twenty-eight vectors represents the portion of the vibration diagram for one fundamental period of

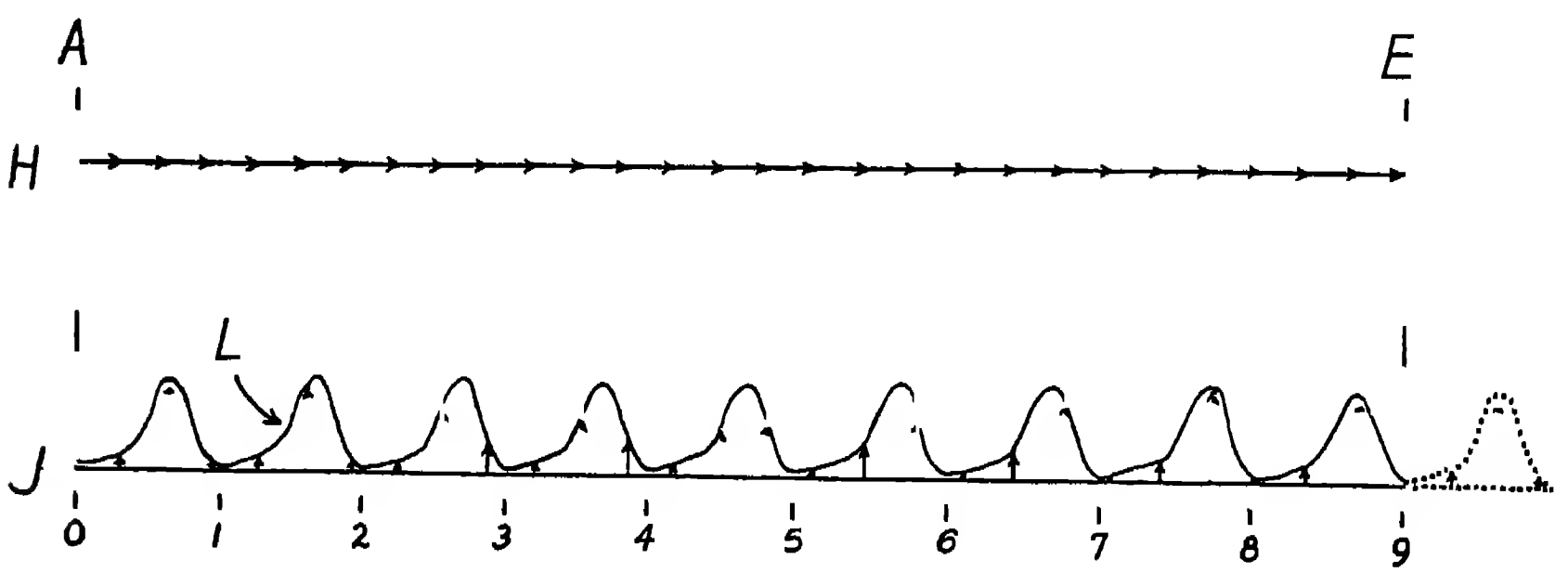


FIG. 156

the error for an ideal grating corresponding to our actual one, at one of the principal maxima. The effect of the errors of position of the grooves may be represented by perpendicular arrows (see diagram J). (The lengths of the perpendicular arrows are much exaggerated.) Let us suppose that the ideal grating is so placed with reference to the actual grating that all grooves having error of position will lie to the *right* of their correct positions. Then, following previously used conventions, the perpendicular arrows will all point upward, as illustrated. The variation in the strain on the ruling machine due to the non-uniformity of the belt will repeat itself every revolution of the belt (see curve L). The error of position of any line will sup-

posedly be proportional to the added strain which existed at the time the line was ruled. In accordance with this supposition the heads of the perpendicular arrows are represented as falling always on the curve L. Now it is to be noted that the errors of position, measured by the lengths of the arrows, show a *tendency* to repeat themselves with each revolution of the belt, but they repeat themselves *exactly* only every nine revolutions of the belt. Therefore the fundamental period of the error is $N=9\times 3\ 1/3=28$.

Let us now reflect upon what happens as we leave the center of a principal maximum. As the vibration curve begins to wind up, the perpendicular vectors will all point inward. For reasons explained in discussing Rowland ghosts, there can be ghosts only for integral numbers of convolutions of the vibration curve per fundamental period of the error—and there will not always be ghosts then. When the curve makes one convolution per fundamental period of the error, the perpendicular vectors will be distributed almost symmetrically around the circle. Their sum will be practically zero. For the same reason the sum will be practically zero for any low number of convolutions per period of the error. When, however, the curve makes a number of convolutions which, in our case, is close to nine, then the sum may attain a value sensibly different from zero, that is, there may be a ghost; and for nine convolutions the sum will be a maximum—we shall then have the strongest ghost.

The foregoing discussion is intended to convey a general idea of how Lyman ghosts come into being. However, the reader is not to suppose that these ghosts always appear in groups of three. The number appearing is a question of intensity; only the strongest one may be intense enough to be recorded. Moreover, the groups, however constituted, do not always fall near the points 1/3 of the distance between principal maxima. They fell near these points for the gratings studied by Lyman, but for the gratings ruled by Anderson on Rowland's 15,000 machine, the groups center about 1/5, 2/5, etc. (In ruling the gratings the pulley made very approximately 5 revolutions for each 2 revolutions of the belt.)

The following remarks may be helpful to the spectroscopist who fears mistaking Lyman ghosts for real spectrum lines: The intensity of the ghosts seldom exceeds 1/10,000 of the intensity of the parent-line, hence the danger of recording them is ordinarily not great. However, there is great danger whenever there is scant energy in the spectral region being recorded, considered, for example, in photographic work from the point of view of the required length of exposure, and when, at the same time, there is much energy in wave-lengths which are not wanted but are nevertheless permitted to enter the spectrograph. (The parent-line may fall far away from the photographic plate or other recording device and the ghosts fall upon it.)

The situation is particularly dangerous in photographing in the infra-red or the extreme ultra-violet. Visual exploration of a spectrum having a few strong lines will reveal those Lyman ghosts which are most to be feared because of their intensity. A strong green parent-line, for example, may give rise to a faint green line falling among red spectrum lines; the faint green line appears in a region of the spectrum where it does not at all belong. By determining visually, upon a circle spectrometer or otherwise, the apparent wave-length, λ', of the ghost, and knowing the wave-length λ, of the parent-line, a constant $\lambda'/\lambda = c_1$ may be determined. Other ghosts of the same parent-line, if found, will lead to constants c_2, c_3, etc. The constants so found are to be applied, when photographing, to the most intense lines of the spectrum (regardless of whether they fall on the plate or not), intensity being judged from the point of view of exposure time. The corresponding ghosts, if present upon the photographs, are then readily identified. If these ghosts are quite faint, as they usually are, there is little danger from additional, fainter Lyman ghosts. A more thoroughgoing method of exploration is to admit to the spectrograph only a single wave-length (or a close group of lines), and then, making long exposures, photograph throughout the intervening region between adjacent principal maxima.

We shall close with some remarks of a historical nature, at the same time giving additional references: Lyman, upon discovering the ghosts in 1901, sent his photographic plates to C. Runge, who thereupon ingeniously hypothecated a type of error which would account for them. Runge's theory was outlined in Lyman's second paper. Then the study of Lyman ghosts rested, at least in so far as published accounts are concerned, for about a score of years. In 1922 Meggers and Kiess,* having been troubled by the presence of ghosts of this type in their photographic explorations of various spectra in the infra-red, made a detailed report regarding the Lyman ghosts shown by two of their gratings. (Two other gratings failed to show them.) They had communicated their results in advance to Runge and to Anderson, who each joined a discussion to the report (see footnote). Anderson contributed the important information connecting the source of the periodic error with the driving belt of the ruling machine. A year later Runge† published a mathematically elaborated presentation of his theory (still based on the same type of error which he had originally hypothecated). In 1924 R. W. Wood‡ gave an account of some interesting experiments, and some

* "False Spectra from Diffraction Gratings," *JOSA and RSI*, **6**, 417, 1922: Part I, "Secondary Spectra," by W. F. Meggers and C. C. Kiess; Part II, "Theory of Lyman Ghosts," by Carl Runge; Part III, "Periodic Errors in Ruling Machines," by J. A. Anderson.

† *Annalen d. Phys.* (4), **71**, 178, 1923.

‡ *Phil. Mag.* (6), **48**, 497, 1924.

observations regarding the driving belt. He also introduced an elastic drive designed to eliminate the appearance of the ghosts.

The explanation of Lyman ghosts which we have given is related to Runge's explanation, but modifies the supposed type of error in the light of later developments.

39. Ghosts Very Close to the Parent-Line.—Very faint ghosts are occasionally found between the parent-line and the first-order Rowland ghost. They are most likely to be found with grating spectrographs of high dispersion and resolving power. There is danger of confusing these ghosts with real hyperfine structure. They have upon occasion been attributed to error which is accidental yet does not follow a Gaussian error curve closely, but it is doubtful whether this be the correct explanation. (Accidental error will be discussed later.) The ghosts in question have recently been studied by R. W. Wood,* who in an abstracted report of his study attributes them to nonperiodic error of spacing arising from variation of friction in the ruling machine. A detailed report has not as yet been published.

40. Error of Run.—This is an error which is progressive across the grating; it must be very small if the grating is to be fit for use, and may consist either in variation of spacing or in variation of groove form. Progressive variation of spacing may arise from the fact that the pitch of the screw of the ruling machine varies slightly along its length, in which case the screw itself is said to have error of run. Again, this error is introduced when, by accident, variation of temperature occurs during the process of ruling. A more frequent cause is variation of friction, resulting in variation of elastic deformation. Variation of groove form may arise from gradual wear of the diamond, but such wear seldom occurs to a marked degree. The discussion of this section will be limited to variation of spacing.

An ideal comparison grating will be chosen which has the same spacing as the actual grating in the vicinity of the center, with the two gratings coinciding groove for groove in this vicinity.

There are two principal types of error of run and these will now be considered:

In the *first* type the spacing increases (or decreases) all the way across the grating. Referring to Figure 157, the equally spaced circles at the top of the figure represent the ideal grating and the dots below represent the actual grating. Referred to this ideal grating, the lines of the actual grating develop error of position as we pass to either side, the error increasing progressively until we reach the edge of the grating where it amounts in the illustration to one-quarter of a grating space. This is the accumulated error in ruling many thousand lines; it is obvious that the variation of the grating space

* *Phys. Rev.*, **40**, 1038, 1932.

must be very small indeed. It is to be noted that, with error of this type, the *mean spacing* taken across the entire grating is equal to the spacing of the ideal grating.

Let us suppose that the incidence is normal and consider the line AB to represent a diffracted wave front for the principal maximum of order m and wave-length λ, for the ideal grating. The corresponding vibration polygon

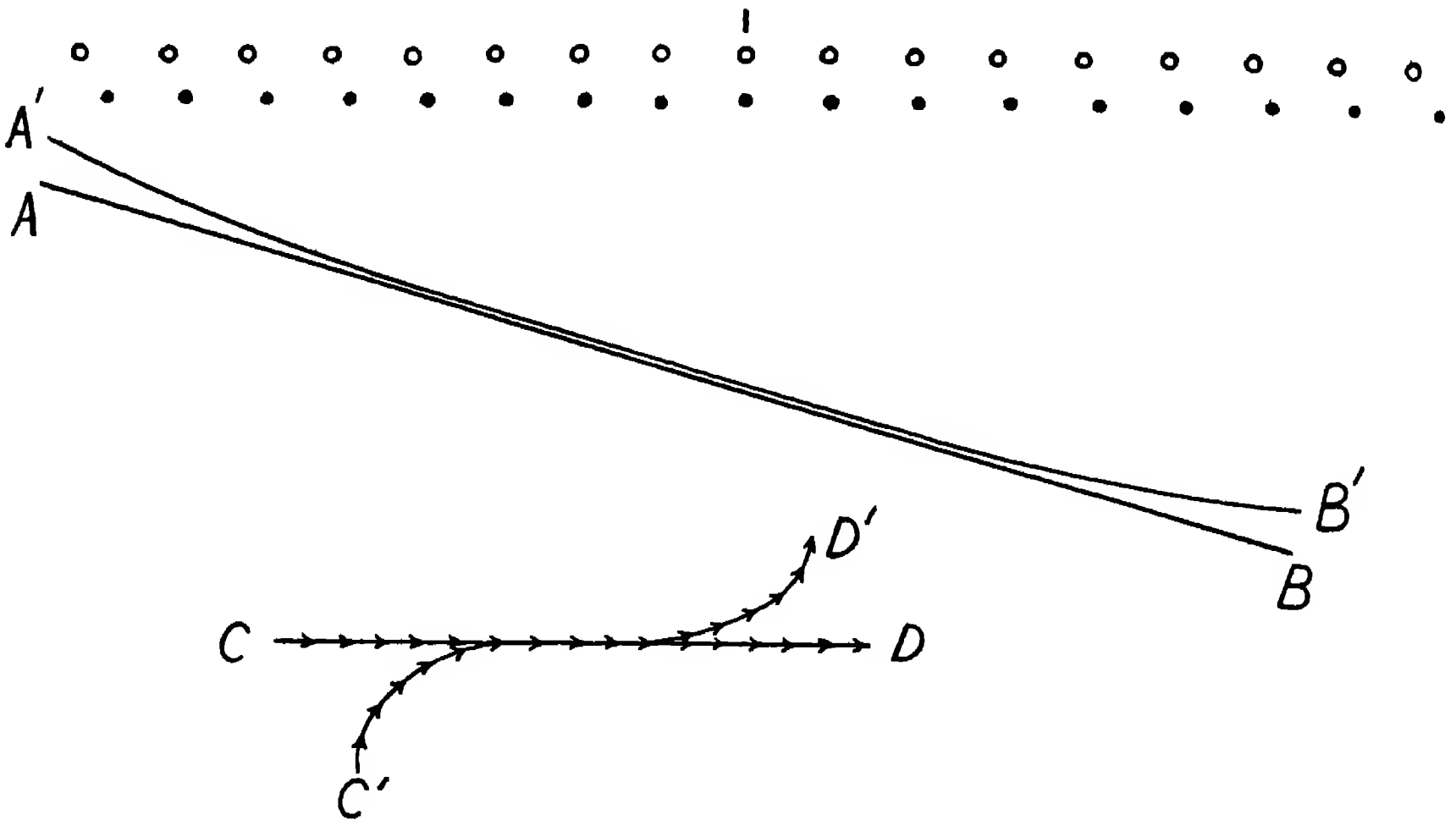


FIG. 157

is CD. For the actual grating the vibration polygon will take the general form of $C'D'$. Moreover, when $m=1$ it will take the particular form illustrated, because then, when the error at the edge is $a/4$, the ends of the vibration polygon, or curve, must make an angle of $\pi/2$ with the central element, as illustrated. For $m=2$, 3, etc., the ends of the vibration curve would make respectively angles of π, $3\pi/2$, etc., with the center—the effect of the error increases with spectral order.

The diffracted wave front, for the actual grating, represented by $A'B'$ is slightly convex. When $m=1$ the distance AA' or BB' is $\lambda/4$. The wave front $A'B'$ will not focus accurately in the principal focal plane of the focusing lens and there will hence be a loss of definition—in this plane. However, by refocusing, by moving, under the illustrated circumstances, to a point somewhat farther from the lens, the lost definition may be regained, either wholly or in part. If $A'B'$ forms part of an accurately circular cylinder, the lost definition may be completely regained. Then the only remaining effect of the error will be a small amount of astigmatism arising from the fact that a vertical cross-section through the diffracted wave front remains a straight line. That is, the horizontal focus remains in the principal focal

plane of the lens; the vertical focus alone moves farther out. (It will be recalled that curvature of the grooves also causes slight astigmatism.) If $A'B'$ is other than accurately circular, much of the lost definition may in general still be regained, but not all of it. The precise form of $A'B'$ is determined by the precise manner in which the error of position of the grooves varies as we pass from the center of the grating toward either edge. Something more will be said upon this subject later.

For diffracted beams on the right of the central image, the focus will be displaced toward the lens.

The following characteristics of this first type of error of run are to be noted: The lines of the grating are spaced unsymmetrically with reference to the central line, but the error of position at equal distances on either side of the center is the same in both magnitude and algebraic sign, the lines lying, in our figure, always on the right of their correct positions. Counting from the center, to the right as positive, a line of running number ν on the right has the same error of position in magnitude and sign as the line of running number $-\nu$ on the left of the center. Hence, when the error of position, E, is represented by a power series in ν, only even powers of ν appear, that is, $E = c\nu^2 + e\nu^4 + \ldots$, where c, e, etc., are constants which may, in general, be either positive or negative. The foregoing facts distinguish what we are designating as the first type of error of run. Further, the vibration polygon for the actual grating has point symmetry about the center, this referring to the polygon corresponding to the position of the maximum for the ideal grating. Hence the pattern for the actual grating is symmetrical about this point and the center is not displaced from its correct position. But, depending upon the magnitude of the error at the edge of the grating, the center may be a minimum instead of a maximum. The pattern is symmetrical in every field plane, including, in particular, the principal focal plane of the lens and the plane of best focus.

Additional facts, unimportant for the general reader, are as follows:

Speaking in terms of the representation of the error E by a power series, when the constant c alone has a value, the diffracted wave front will be circular (the reason for this will become clear upon referring to eq. 5 of chap. 2, which applies to any wave front of circular cross-section; our present ν is proportional to q in this equation, and E is proportional to δ). It follows that the vibration curve for any point in, e.g., the principal focal plane of the lens will be a section of a Cornu spiral; the pattern may be derived by sliding a constant length of arc along the spiral, in close analogy with the method of deriving the Fresnel pattern for a slit. Finally, higher terms of the error, terms in ν^4, ν^6, etc., lead each to a wave front which is not circular and hence has variable curvature. Each such term leads to a spiral of higher order than the Cornu spiral; or rather, since the curvature of the wave front varies from point to point, each point in the pattern calls for one of a *family* of spirals, and there is a

different family for each term of the error. The problem becomes involved but it is still possible to derive the pattern (for details see reference given farther on).

In the *second* principal type of error the spacing increases (or decreases) in passing from the center toward either edge of the grating (see Fig. 158). This type represents in numerous obvious respects a counterpart to the first type. Taking the error of position at the edge of the grating again as $a/4$, and $m=1$, the error of path at the edge of the grating is $\lambda/4$ and the vibration polygon or curve for the actual grating will hence be $C'D'$. This is the curve for the point P which represents the position of the maximum for the ideal grating. (The line F represents the principal focal plane of the focusing lens L.) The curve $C'D'$ is symmetrical about a perpendicular erected at its center, as are likewise other vibration curves which will be referred to later. The diffracted wave front is $A'B'$ and it is obvious that any wave front having this general form will not come to an accurate focus either in the principal focal plane or elsewhere; refocusing will not improve the situation. Hence this form of error constitutes a serious fault. It is damaging in the important respect of definition and resolving power.

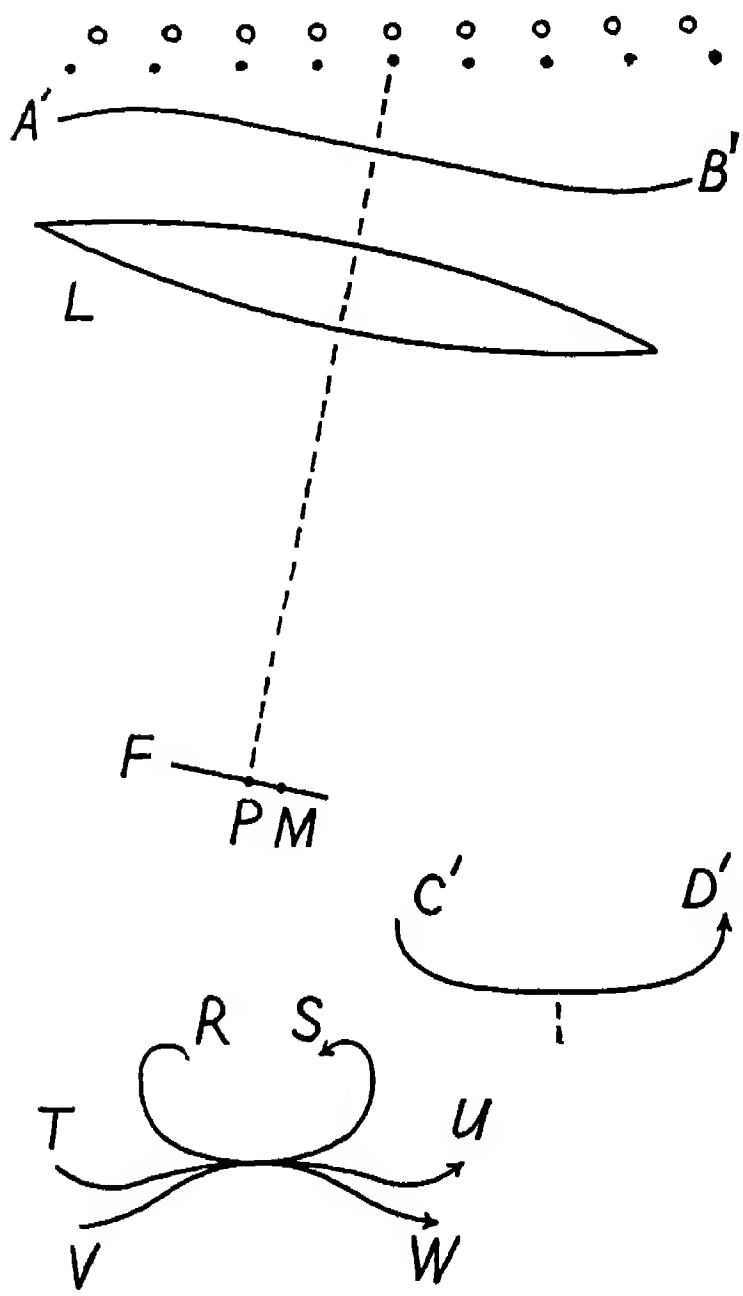


FIG. 158

It is to be noted that the mean spacing of the actual grating is greater, in our figure, than that of the ideal grating which we are considering. Consequently, an ideal grating having a spacing equal to the mean spacing of our actual grating would have the corresponding principal maximum slightly closer to the central image, let us say at M.

By considering the actual grating by portions—by a central portion and two outlying portions—and considering corresponding portions of the diffracted wave front, it is easily deduced that the maximum of the pattern falls to the right of P, and that the pattern is not symmetrical. It is possible to derive the form of the pattern by following the vibration curve through various stages. Suffice it, however, to point out that as we leave P and pass to the left, the curve curls up yielding, for example, RS. In passing from P

to the right the curve uncurls, the closing vector at first increasing in length. The closing vector passes through a maximum, curve TU, representing the maximum of illumination on the right of P, which occurs before reaching M. At a certain later stage we have curve VW which has horizontal ends. This stage is for the point M in the pattern where the maximum would lie for an ideal grating having the same mean spacing as the actual one. In comparison with this ideal grating the grooves at the edges would be in their correct positions. Referring now to Figure 159, the curve drawn in broken line represents the pattern for an ideal grating having the same grating space as our actual grating at the center. The center of this curve is the position which we have designated by P. The solid curve is for an actual grating when the maximum error of path is $\lambda/4$ (the case which we have been supposing). To the right of the maximum of the solid curve, M indicates the position of the maximum for an ideal grating having the same mean spacing as our actual one. The dotted curve and the position M' have corresponding significance when the extreme path error is $\lambda/2$.

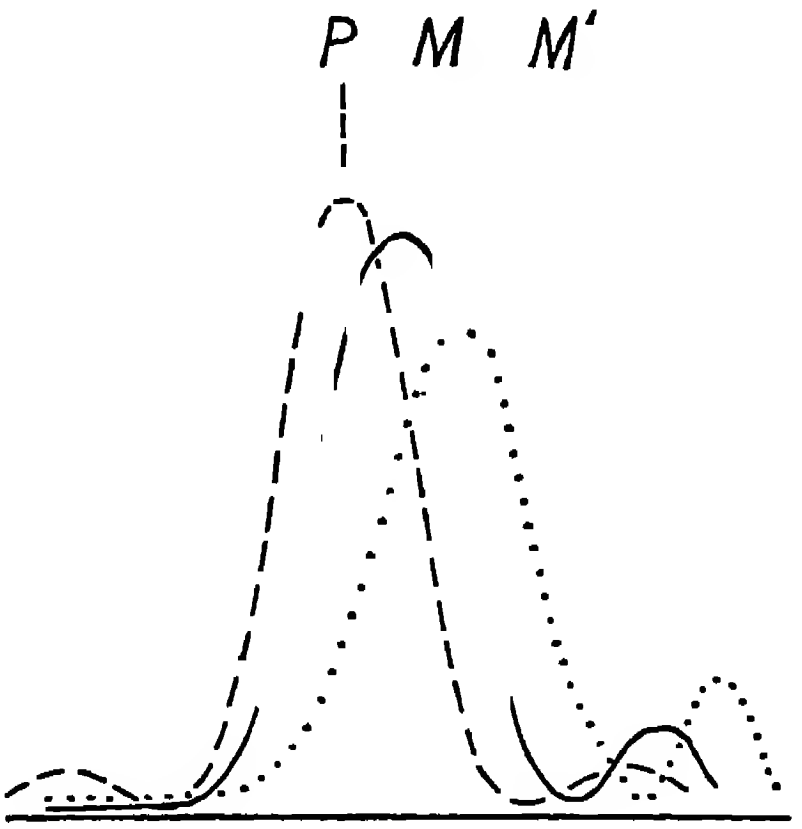


FIG. 159

The foregoing, second type of error of run is characterized by the fact that the error when represented by a power series involves only terms having odd exponents of ν. The term in ν^3 usually dominates the series. This term is really the first term of the series; when the ideal grating is chosen to have the same spacing as the actual grating in the vicinity of the point where ν equals zero, then the term in ν to the first power drops out.* The curves of Figure 159 are drawn on the supposition that the error is represented by the term in ν^3 alone.

Any actual error of run may be represented by a power series having both odd and even terms. The even terms represent an error of the first type and the odd terms represent one of the second type. Any actual error of run can thus be resolved into two errors, one of these being of each type.

* We may see this by considering two ideal gratings, I_1 and I_2, having different spacings, and referring I_2 to I_1 as a standard. The "error" of position of the groove of number ν of the grating I_2 then becomes $E = b\nu$, where b is a constant. That is, the first-power term arises from choosing as an ideal grating one having, so to say, the wrong spacing, and this term drops out when the correct spacing is chosen.

The first type and the second type correspond respectively to aberrations of even order and odd order of optical systems in general.

The reader who desires further information is referred to the excellent article, "Theory of Imperfect Gratings," by C. M. Sparrow.*

The explanation of the focal peculiarities of gratings was first given by A. Cornu.† Rowland also treated the subject briefly.‡

41. Accidental Error.—The errors thus far considered have been systematic. The error now to be considered is distinguished from these by the fact that the error of position of the grooves (when error of spacing is in question) follows the so-called "law of error"—the law of accidental error. The groove of running number ν may be supposed to have an accidental error of position E which is a small fraction of the grating space a. In the direction of the central image the vector polygon is a straight horizontal line of n vectors of equal length, n being the total number of lines of the grating. In the direction of a principal maximum other than the central image, the polygon is a slightly zigzag line, the angle which any given vector makes with the horizontal being $2\pi m(E/a)$, where m is the spectral order. The effect of the error increases with spectral order. We can resolve each vector along and at right angles to the direction it would have for an ideal comparison grating of the same mean spacing. Since E is small compared to a, the component along this direction will have practically the same length as for the ideal grating (as explained in detail in the latter half of sec. 37) and we need consider only the effect of the short perpendicular vectors.

Let us consider their effect at some point far removed from a principal maximum—for example, at the position of the pth minimum for the ideal grating. The vibration polygon for the ideal grating now makes p convolutions, or sweeps out an angle of $2\pi p$; and the successive vectors of the polygon make small equal angles of $2\pi p/n$ with each other, n being the total number of lines of the grating. The small perpendicular vectors may now be placed all with their points of origin at a common point. They will radiate out from this common point, each making an angle of $2\pi p/n$ with the next. Their lengths are purely accidental and positive and negative values are equally probable. Their sum will however not in general be zero. There

* *Astrophys. Jour.*, **49**, 65, 1919. In this article the theory is developed in somewhat different terms than those used above. The derivative of the error of position with reference to the groove number, i.e., $dE/d\nu$ in our terms, is proportional to the curvature of the vibration curve due to the error. The theory is developed in terms of curvature of the vibration curve. As the result of the differentiation involved, the errors designated above as those of the second order, the third order, etc., become respectively those of the first order, the second order, etc.

† *Comptes rendus*, **80**, 645, 1875; **116**, 1215, 1893.

‡ *Op. cit.*, p. 542.

is a certain small "expectation" for the magnitude of this sum, which is determined by the law of error as applied to this case.* Hence there will be a small amount of light at the position of what would be the minimum for the ideal grating in question. The position of a minimum was chosen merely for convenience of consideration. The situation is much the same at other points of the field plane, as the reader may conclude for himself. The general conclusion is that a wave-length λ will send a small amount of light to all parts of the field plane. That is, accidental error causes scattered light. The amount of this is small for any one spectrum line of ordinary intensity, but when contributed to by the numerous lines of a given spectrum it may be appreciable. The intensity of the scattered light due to accidental error of spacing is zero in the vicinity of the central image and increases with the square of the spectral order.

Accidental variation of groove *form* may be caused by the crystalline structure of the speculum plate upon which the grating is ruled. This leads, at each of the principal maxima including the central image, to a vibration polygon consisting of vectors of slightly unequal *length*, however all lying accurately in a straight line. By considering each vector to be the sum of a "mean length," plus or minus an "error," it is readily deduced that accidental variation of groove form will also produce scattered light. However, the amount will now be independent of spectral order; there will be scattered light also in the vicinity of the central image.

42. Abrupt Error.—This is neither systematic nor accidental error. The effect of a single abrupt error of spacing occurring at the center of the grating has been considered by Michelson.† Each half of the grating is supposedly ideal and both halves have the same spacing a, but at the center of the grating there is one space differing markedly from a. Sparrow, in the aforementioned article (p. 79), gives the vector treatment for this case. There is some question about the actual occurrence of such error; in any event it is not one of great practical importance. The effect of the error is very similar to that of error of run of the second type.

Abrupt variation of groove form may be caused by breakage of the ruling point. This case is also discussed by Sparrow (on p. 85 of his article); it is not a case of great importance. Suffice it to call attention to the fact that a breakage resulting in a loss of light from all of the grooves ruled subsequently is equivalent to screening off a proper length of the ends of the

* For details see p. 82 of the article by C. M. Sparrow referred to at the close of the previous section.

For a mathematical treatment of the effect of accidental error see also E. Buchwald, *Annalen d. Phys.*, **80**, 279, 1926.

† *Astrophys. Jour.*, **18**, 278, 1903.

grooves of an ideal grating, from the point where the breakage occurred, for the remainder of the grating.

43. Variation of Groove Form.—This is the third of the primary classes of error designated at the beginning of the discussion of error of ruling (sec. 34). Accidental variation of groove form and abrupt variation have just been touched upon. Theoretically, to each type of spacing error (otherwise known as phase error) there corresponds a type of variation of groove form (otherwise known as amplitude error). Actually, amplitude error is of much less importance than spacing error.

The analogue of periodic error of spacing does not occur. However, it may be of interest to point out that if periodic amplitude error did occur it would give rise to ghosts in the same positions at which they occur due to corresponding spacing error. The reader may readily deduce this fact. The principal difference would be that, for amplitude error, the intensity of the ghosts would be independent of spectral order, and accordingly, there would also be ghosts in the vicinity of the central image.

Amplitude error of run, arising from gradual wear of the diamond, is the analogue of spacing error of run of the first type. The case is briefly discussed by Sparrow (p. 83 of his article). While wear of the diamond undoubtedly occurs, it develops that the effect is negligible for any reasonably assumed value of the variation of the light contributed by the grooves. The analogue of the second type of spacing error of run would not be expected to occur.

The effects produced by amplitude error of a given type bear a general resemblance to the effects produced by the corresponding type of spacing error. For example, periodic error would in both cases produce ghosts, and accidental error in both cases produces scattered light. There is, however, always this difference: Whereas the effects of phase error are absent at the central image, and increase with spectral order, the effects of amplitude error are revealed with equal relative intensity in the vicinity of the central image and the various spectral orders.

CHAPTER 7

SUNDRY RAMIFICATIONS OF DIFFRACTION THEORY

In this chapter we shall consider several applications of diffraction theory to the theory of optical instruments, and to certain optical methods of measurement. We shall also consider several hitherto barely mentioned general aspects of diffraction theory without direct regard to their applicability. Further, there will be some matters treated which do not properly fall under the heading of diffraction at all, in order to prepare the way for the discussion of other matters which do properly fall into this field.

1. Permitted Aberration.—To say of an optical system, or, for example, of a single lens, that it is perfect with regard to given conjugate object and image points, means that the various optical paths from the object point by way of various portions of the lens to the image point are all equal. The image will then be as nearly perfect as the wave nature of light permits it to be, and it is only because the wave-length of light is as short as it is that we are able to obtain as highly perfect images as we do obtain.

The question arises: Within what limits must the aberration or error of path be held if it is to cause no marked impairment of the most perfect image which we can hope to form with an optical system of given general characteristics? This question was raised and fairly well answered in 1859 by L. Foucault.* The answer was later given more accurately and with more complete explanation by Lord Rayleigh.† It may be stated thus: If an image formed by any optical system is to be sensibly perfect the error of optical path should nowhere greatly exceed $\lambda/4$.

We shall refer to the last two figures of the preceding chapter—namely, to Figures 158 and 159—which were previously considered as applying to a grating having error of run, the error of optical path being supposedly introduced by error of position of the grooves. From our present point of view the cause of the error of path is immaterial and the two figures in question may be supposed to apply to any optical system. However, referring first to Figure 158, we shall suppose that the error is introduced before the beam passes through the final, focusing lens L, and we shall consider this lens to be perfect. The vibration curve $C'D'$ is drawn on the supposition that the wave front $A'B'$ deviates from being plane by an amount $\lambda/4$ at each edge of the aperture. That is, the error of path at the edge is supposedly $\lambda/4$.

* *Ann. de l'Observ. Imperial de Paris*, **5**, 197, 1859.

† *Scientific Papers*, **1**, 428 and 436.

If the aperture of the system were extended, widened, with a corresponding increase in the maximum error, the contributions from the added portions would add sections to the vibration curve which would cause the closing vector of this curve to be shorter, resulting in a decrease of intensity. However, the ensuing decrease of intensity would not in itself be the most damaging feature of the situation. Since the total light which goes into the image is determined by the size of the aperture and is unaffected by the error, the light which is lost at one part of the image spreads into other parts—the image broadens and, with the particular type of error under consideration, a cubic error, a lateral maximum develops on one side of the principal maximum (see Fig. 159). Here the solid curve is for a maximum path error of $\lambda/4$ (in accordance with the previous figure). The image is slightly impaired, but only to an amount which can be tolerated. But for a path error of $\lambda/2$ at the edge of the aperture, dotted curve, the image has broadened appreciably and the lateral maximum has attained damaging intensity. We can no longer hope to attain very near to the definition and resolving power of which the system is theoretically capable.

A path error of $\lambda/4$ allows a mirror to have an error of only $\lambda/8$ at normal incidence, because an error of figure of this amount will introduce a path error of $\lambda/8$ at incidence and an equal error upon emergence. On the other hand, with a lens, if ϵ be the error of figure and μ the index of the glass, we have $\epsilon(\mu-1)=\lambda/4$. Since for glass $(\mu-1)=1/2$ roughly, we have $\epsilon=\lambda/2$. That is, the allowable error of figure for a lens is about four times as great as for a mirror.

There is perhaps some danger that the foregoing principle may be misinterpreted. The error of path may exceed $\lambda/4$ when there is a possibility of compensating by refocusing (see chap. 6, sec. 40). Further, aside from this possibility, an optical system does not, for all uses, need to be held to an error of path of $\lambda/4$. For example, a lens may deviate considerably from correct figure without impairment of its usefulness as a reading lens or as a condensing lens. The reasons for this the reader will no doubt divine.

2. Resolving Power of the Telescope.—Foucault also ascertained the conditions which determine whether two small objects close together will be "resolved" by a telescope or whether they will be seen as a single object. Again we are indebted to Lord Rayleigh* for clarifying discussion.

Let us suppose that the objective lens of the telescope is perfect and that the images are consequently as perfect as the nature of light allows. Referring to Figure 160, let O represent a point source of monochromatic light of wave-length λ, and L the objective of the telescope. Ordinarily the beam entering a telescope is limited only by the aperture of the lens itself, but we

* *Ibid.*, pp. 418 and 488; **2**, 410, etc.

shall suppose for the sake of discussion that a diaphragm, DD, is situated close in front of the objective; there is no loss of generality in supposing this. Let MN represent a wave front incident from O. This is brought to a focus in the plane PP'. The image is not a mathematical point but is a diffraction pattern p. This image is viewed through the eyepiece of the telescope represented by e.

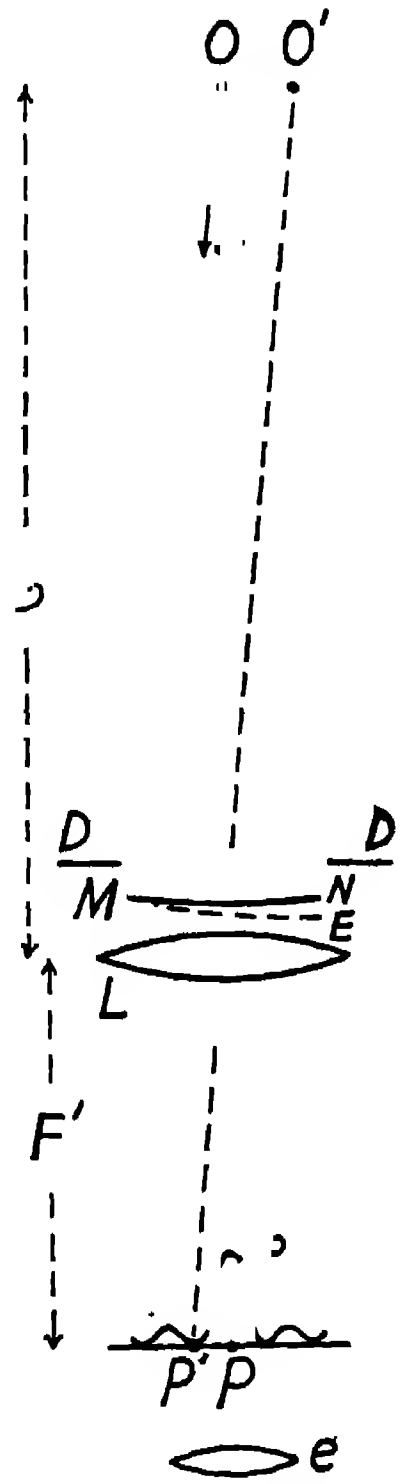


FIG. 160

Suppose O' is a point source similar to O. The image of O' will be a diffraction pattern similar to p but displaced to the left. Suppose, to begin with, that O' lies farther to the right of O than illustrated. Then the two diffraction patterns will be separate as shown by p and p' of Figure 161a, and thus the images will be seen separately. If now O' approaches O, the pattern p' approaches p and when the central maximum of each pattern lies over the first minimum of the other pattern we have the condition represented in Figure 161b. The resultant illumination is obtained by summing up the two intensity curves and is represented by the dotted curve, which, it will be observed, has a slight minimum at the center but only a slight one. A finite fall in the intensity, between the two maxima, is required if the images are to be resolved. If O' approaches O more closely, the minimum will rapidly disappear and soon be converted into a maximum. Accordingly, the condition that each maximum should fall on the first minimum of the other pattern is regarded as the criterion which establishes the limit of resolution. If the maxima lie at this distance or farther apart, they will be seen separately; if they lie closer together, they will be seen as one image. However, there is a certain degree of arbitrariness in selecting exactly the point where each maximum falls on the first minimum of the other pattern.

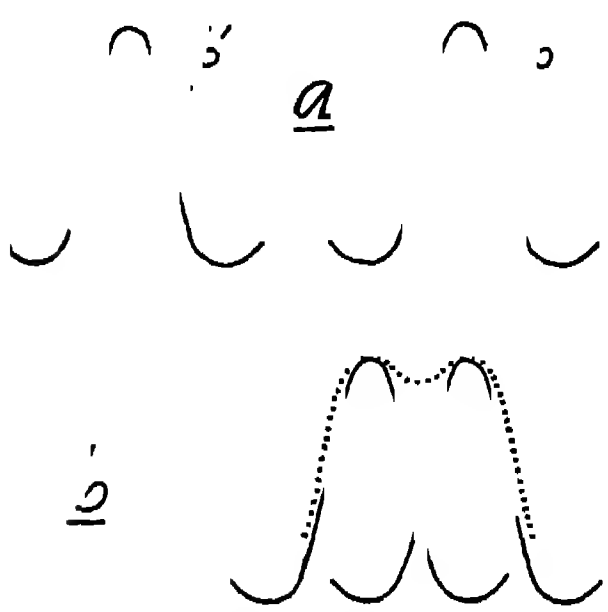


FIG. 161

If the aperture in the diaphragm D is rectangular, the first minimum in the diffraction pattern p of Figure 160 will come when the distance NE between the right-hand edges of the actual wave front MN and the diffracted wave front ME is equal to λ. The angle between these two wave

fronts is then λ/w, where w is the width of the aperture. If the maximum of the pattern due to O' is to fall on the first minimum of the pattern due to O, then ME must be an actual wave front for O'. That is, the actual wave fronts from O and O' must make an angle λ/w with each other. The angle which these wave fronts make with each other is also the angle subtended by the sources O and O' at the telescope objective (incidentally, this is furthermore the angle subtended by the images at the objective). Denoting the distance from O, or O', to the objective by b, and the distance from the objective to the plane in which the images are formed by F', we have:

$$\frac{OO'}{b}=\frac{\lambda}{w}, \qquad \left(=\frac{PP'}{F'}\right). \tag{1}$$

When the distance b is large, as it usually is, then the distance F' is very nearly equal to the focal length of the objective. The foregoing equation tells us that the theoretical least angle or resolution is equal to the wavelength divided by the width of the aperture.

The verification of equation (1) presents an instructive laboratory exercise. A rectangular aperture of variable width should be used, the sources O and O' being conveniently maintained at a fixed separation and at a fixed distance from the telescope. Point sources may to advantage be replaced by narrow vertical slits or, better still, by a grid having a series of narrow slits in it, the grid being backed by a sodium flame. The grid may be obtained by ruling lines upon a chart and photographing this to reduced scale.

If the aperture limiting the beam is circular, then the first minimum in the diffraction pattern does not occur until the distance $NE=1.22\lambda$ (see chap. 5, sec. 3). Therefore, for a circular aperture:

$$\frac{OO'}{b}=\frac{1.22\lambda}{d}, \tag{2}$$

where d is the diameter of the aperture. The verification of this equation may also be carried out as a laboratory exercise, but it is less convenient than the verification for a rectangular aperture.*

The *resolving power* of a telescope is defined as the reciprocal of the least angle of resolution. However, the least angle itself is often stated, in place

* There are several reasons for this: A circular aperture cannot be varied continuously in diameter. The theoretical resolving power is usually exceeded by 10 per cent or more—a fact which requires explanation: At the theoretical limit for a circular aperture the minimum is slightly more pronounced than is required for resolution—more pronounced than at the theoretical limit for a rectangular aperture. Also, when a circular aperture is used, it is not quite permissible to resort to the convenience of replacing the point sources by slits (see Rayleigh, *ibid.*, **1**, 418, and **3**, 95).

of the reciprocal, this angle being smaller in direct proportion as the resolving power is larger. Equations (1) and (2) respectively determine the so-called *theoretical* resolving power for telescopes provided with rectangular and with circular apertures. We cannot hope to attain the resolving power determined by these equations if the objective lens is markedly imperfect or when the instrument is imperfectly focused. However, in practice the theoretical resolving power is attained by a good telescope. There is, moreover, no occasion for surprise when a given set of observations indicates that the theoretical resolving power has been slightly surpassed, for, as has been pointed out, there is a certain degree of arbitrariness in the establishment of what we call the "theoretical limit."

It is to be noted that the resolving power of a telescope is independent of the focal length, and also that it increases as the wave-length of the light decreases.

Supposing that a telescope objective resolves two object points, it is further necessary that an eyepiece of sufficient magnification be used in order to see the images separately. The least angle of resolution for the normal human eye is about one and one-half minutes of arc. Accordingly, the eyepiece must be one of sufficient power to cause the beams from the two images formed in the focal plane of the objective to leave the eyepiece and enter the eye at an angle of at least one and one-half minutes of arc.

When two object points are to be resolved photographically, the grain of the photographic plate must be sufficiently fine to permit the formation of two separate images at the separation which the images have in the focal plane. In other words, the resolving power of the photographic plate must be sufficiently great.

The power of a telescope to resolve double stars of approximately equal intensity may be somewhat increased by placing a circular stop of, let us say, one-fourth of the diameter of the object glass centrally over this glass. The effect of the stop is to reduce the radius of the central diffraction disk. Such a stop was at one time frequently used. However, there results an increase in the intensity of the first bright ring surrounding the central diffraction disk, and this is a disadvantage in general observation and also in observing double stars of markedly unequal brightness.

3. Resolving Power of the Grating.—We shall suppose that a parallel beam is incident normally upon a plane grating. The theory applies, however, at any angle of incidence and applies to either a plane or a concave grating. Referring to Figure 162, suppose MB is the grating and MN is the diffracted wave front for the principal maximum of order m and wave-length λ. The distance BN is then equal to $mn\lambda$, where n is the total number of lines of the grating. There arises the diffraction pattern p in the focal plane of the

focusing lens L. The first minimum in this pattern will occur in the direction for which the diffracted wave front is ME, the angle which this wave front makes with MN being determined by the fact that the distance NE must be equal to λ. Hence $BE=(mn+1)\lambda$. Further, a wave-length greater than λ by an amount $\Delta\lambda$, namely, $\lambda'=\lambda+\Delta\lambda$, will give rise to a diffraction pattern lying to the left of p. The path increment at the edge of the grating for λ' will be $mn\lambda'$. The wave-length λ' will just be resolved with reference to λ if the diffraction pattern for λ' has its center coincident with the first minimum of p, as illustrated by p'. The principal diffracted wave front for λ' will then be ME and then $BE=mn\lambda'=mn\times(\lambda+\Delta\lambda)$. Equating this value of BE to the value previously found, we have:

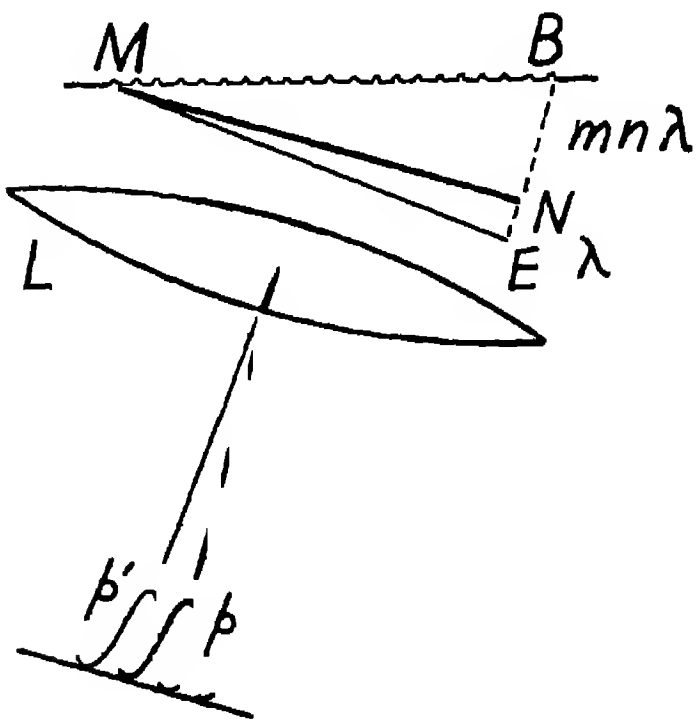


FIG. 162

$$(mn+1)\lambda=mn(\lambda+\Delta\lambda)\,,$$

whence, canceling $mn\lambda$, which is common to the two sides of the equation, and dividing through by $\Delta\lambda$, we have:

$$\frac{\lambda}{\Delta\lambda}=mn\,. \tag{3}$$

The resolving power of a spectroscope or spectrograph of any type is defined as being equal to $\lambda/\Delta\lambda$, where λ is the wave-length of one of two spectrum lines which can just be resolved and $\Delta\lambda$ is the difference in wave-length between the two lines. For example, the two sodium lines have wave-lengths $\lambda=5890$ Å and $\lambda'=5896$ Å. Hence $\Delta\lambda=6$ Å and the power required to resolve them is:

$$\frac{\lambda}{\Delta\lambda}=\frac{5890}{6}=982\,.$$

The value of $\Delta\lambda$ is always small compared to λ. Since resolving power is not a quantity which can with reason be evaluated to a high degree of accuracy, it becomes a matter of indifference as to which of two just resolvable wave-lengths we place in the numerator; thus, in the example above, placing the greater of the two wave-lengths in the numerator, we have $5896/6=983$, which differs inappreciably from the previous value.

It is evident from equation (3) and the example given that a grating of about 980 lines will just resolve the sodium lines in the first-order spectrum and will easily resolve them in all higher orders, and a grating of about 490 lines will resolve the sodium lines in the second and higher orders.

The product mn, the order of the spectrum times the number of lines of the grating, is the theoretical resolving power of the grating. This obtains when there are no imperfections in the optical system. When there are imperfections the value of $\Delta\lambda$ required for resolution will be greater, and $\lambda/\Delta\lambda$, which is always the actual resolving power, will then be smaller than mn.

The theoretical resolving power of a grating is proportional to the total number of lines of the grating and is independent of the width of the grating and the grating space. Or rather, the foregoing statement is the most illuminating brief statement which we can make of what determines the resolving power. However, the question of what determines the value of a given quantity is often a matter of point of view, the value itself being at the same time quite definite. So it is in the present case, as we shall now show. Denote the angle of diffraction by θ (the angle of incidence we shall suppose always equal to zero). Denote the width of the grating by w. Then $n=w/a$, where a is the grating space, and further, according to the grating law, $m=(a \sin \theta)/\lambda$. Referring to equation (3) and to Figure 162, we have:

$$\frac{\lambda}{\Delta\lambda}=mn=\frac{\text{Distance } BN}{\lambda}=\frac{w \sin \theta}{\lambda}. \tag{4}$$

That is, we may regard the resolving power at a given wave-length as being determined solely by the distance BN, and we are at liberty to regard this distance as being determined, among other factors, by the width of the grating and the grating space, of which we previously considered the resolving power as being independent. Again we may regard the resolving power at a given wave-length and angle of diffraction as being determined solely by the width of the grating, and the resolving power is in this sense independent of both the number of lines and the grating space. There is of course no real contradiction between these various points of view. Suppose we have two gratings having the same number of lines but different grating spaces. When we say that the resolving power is independent of the width of the grating and the grating space, we tacitly assume that we shall alter the angle of diffraction as may be necessary. When, on the other hand, we keep the width and the angle of diffraction fixed, then if we double the number of lines we shall halve the spectral order.

4. The Law of Extreme Path.—This law is a generalized formulation of Fermat's *principle of least time*, and is a law of geometrical optics which does not properly fall within our province. It is here taken up only for the

purpose of deriving certain conclusions of which we shall want to make use. The treatment is not a complete one.

When a ray travels a distance d in a medium of index of refraction μ, the optical length of the path in question is by definition equal to μd. Now $\mu = c/v$, where c is the velocity of light in air or vacuum and v is the velocity of light in the medium of index μ. Hence $\mu d = cd/v$, and since $d/v = t$, where t is the time taken, we have $\mu d = ct$. That is, the optical length of any path is the product of the time taken, and the velocity of light in air or vacuum—or, in other words, it is the geometrical distance which the disturbance would travel in the same length of time in air or vacuum. Optical distance and time taken are related by the absolute constant c and hence any theorem which applies to one applies to the other. In case a ray travels through more than one medium, the individual optical distances are to be added, and the same, of course, holds for the individual lengths of time.

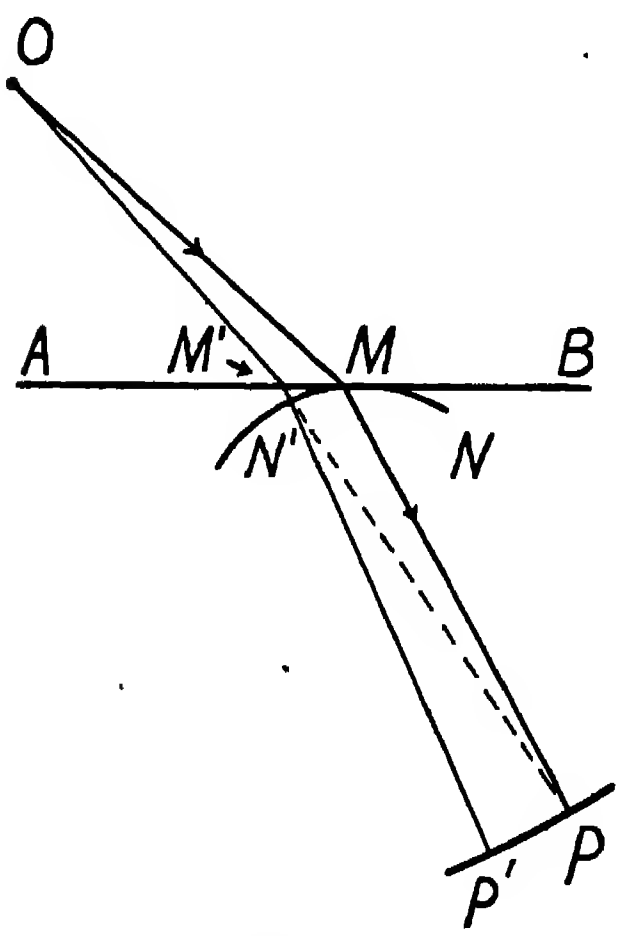


FIG. 163

Referring to Figure 163, suppose O is a point source in air or vacuum, index of refraction equals unity, and AB is the bounding surface of a second medium, let us say glass, of index μ. Let P be an arbitrarily chosen field point in the second medium and let OMP represent the ray reaching this field point. We shall now show that the optical length of the actual path, $OM + \mu MP$, is less than the optical length of any path which is not an actual path, less than, for example, $OM' + \mu M'P$, this holding true whether M' lies as indicated or on the other side of M.

At the point of the surface AB which lies along the actual path, namely, at M, erect a surface $N'N$ such that if it were the bounding surface of the second medium, in place of AB, all rays refracted at $N'N$ would be directed toward the point P. The surface $N'N$ would then be a so-called "aplanatic" surface for the medium of index μ with reference to the conjugate points O and P. The word "aplanatic" is derived from the Greek α, "not," and πλανατικός, "wandering," and is simply descriptive of the fact that the rays from O are to pass accurately through P and not wander from this point. Of course P would then be an image of O and any wave front in the second medium would be a sphere converging about P. Disturbances from various points on a given wave front would all reach P at the same time. That is, the periods of transit from O by way of various points of $N'N$ to P would all

be equal, and hence it follows further that when a surface is aplanatic with reference to two points O and P then the lengths of the optical paths from O by way of various points of the surface to P are all equal.

In view of the foregoing, the optical path $OM+\mu MP$ must be equal to the optical path from O, in air or vacuum to N', and thence in glass to P, and this path is less than the path $OM'+\mu M'P$ (which obtains when AB is the bounding surface) because the last-mentioned path has a greater length in glass than the path by way of N' when $N'N$ is the bounding surface. This holds, whatever the position of N' and M' may be with respect to M. Hence when AB is the bounding surface, the actual path $OM+\mu MP$ is shorter than all other paths. That is, when the bounding surface between two mediums is plane, the actual path is a *minimum*. This is the case which Fermat treated.

Further, the actual path will be a minimum also if the bounding surface under consideration, AB, instead of being plane, is curved, yet still falls outside of the aplanatic surface, $N'N$, everywhere except at M. The point M is in general, by hypothesis, that point of the bounding surface, whatever this may be, which lies along the actual path, and the aplanatic reference surface, $N'N$, is made tangent to the surface under consideration at M, by construction.

When the actual surface is such that it lies *inside* of the aplanatic reference surface, $N'N$, everywhere except at M, it follows by similar reasoning that the actual path is a *maximum*.

Thus the actual path is either a maximum or a minimum—it is an *extreme*. But see also the next paragraph and its accompanying footnote.

There are additional general conclusions to be drawn. Suppose $M'M$ is a *small* distance. The path $OM'P$ exceeds OMP, but because OMP is an extreme value (i.e., a minimum) the increment of optical path by way of $OM'P$ is even *much smaller* than $M'M$. Consider the point M' as moving along AB. Expressed in terms of limits: $\lim_{M'M \doteq 0}$ (increment/$M'M$) $=0$. Geometrically, the extreme smallness of the increment is a result of the fact that the surface under consideration, AB, and the aplanatic surface, $N'N$, are tangent to each other at M. The length of the path $OM'P$ varies as the position of M' varies, but this length is *stationary* for paths in the vicinity of OMP, the actual path.*

* The fact that the path is stationary in the vicinity of the actual path is really the most general statement which can be made, and the name "law of stationary path" would more accurately describe the facts and be more inclusive than "law of extreme path." The latter name appears, however, to have the sanction of usage. While the point corresponding to the actual path must always lie on a horizontal portion of the curve obtained when the path length is plotted against position on the surface, the horizontal portion in question may be neither a maximum nor a minimum, but, in cases which we have

A further conclusion follows: If we are interested in distances down to the order of magnitude of the small distance $M'M$ and are not interested in distances of lower order of magnitude, we may consider various paths in the vicinity of the actual path as all equal. That is, $OM'P$ differs from OMP by much less than MM', and hence if we are not interested in quantities which are much smaller than MM', we may say $OM'P = OMP$. We may arrive at this same conclusion also in another way: Consider the refraction which takes place at the plane surface AB. The rays MP and $M'P'$ are refracted rays, and PP' is a refracted wave front. The optical paths OMP and $OM'P'$ are accurately equal. Now since $M'P'$ is perpendicular to PP', we may replace $M'P'$ by $M'P$ without sensible increase in length. Hence the path $OM'P$ may be considered equal to $OM'P'$ and hence also to OMP—so long as we are not interested in quantities of lower order of magnitude than PP' or MM'.

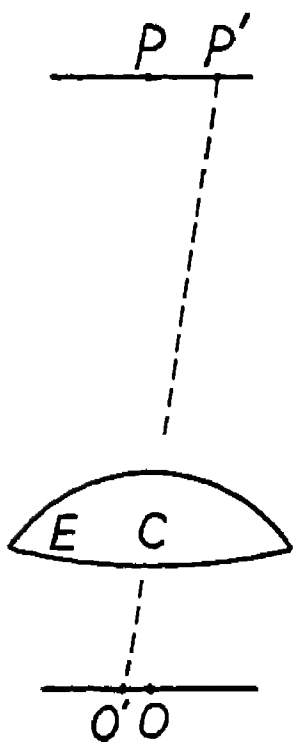


FIG. 164

There is one more conclusion of which we shall want to make use. Referring to Figure 164, suppose O is an object point on the axis of a lens and P is the corresponding image point. Denote the optical distance from O to P by δ. Suppose O' is an object point close to O. Denote the optical distance from O' to P' by δ'. Considering in each case the route by way of the center, C, of the lens, it is obvious, after what we have learned in the foregoing paragraph, that the optical distances δ and δ' differ only by a quantity which is small compared to the small distances OO' or PP'. That is, we may consider δ and δ' equal if we are not interested in distances which are small compared to OO' or PP'. This conclusion holds, however, only when O and O' both lie near the axis of the optical system.

The following relation is of incidental interest: The optical path from O by way of the edge of the lens, E, to P is of course equal to the path by way of the center. That is, $OEP = OCP$. Further, the equal path $O'CP' = O'EP'$. Thus $OEP = O'EP'$, whence $(OE - O'E) = (EP' - EP)$.

5. Resolving Power of Prism.—Referring to Figure 165, suppose ab is a parallel beam from the collimator of the spectroscope, incident upon the prism, P. Let M_0N_0 represent an incident wave front just striking the prism at N_0. This will be a wave front for light of all wave-lengths. In passing through the prism the light is dispersed according to wave-length.

---

not discussed, may be an *inflection*. The word "stationary" is then descriptively correct, but "extreme" is not correct. Again, when the surface under consideration is itself an aplanatic surface, the path length is constant, i.e., "stationary," but not properly "extreme."

Let N_0C represent the ray direction within the prism for light of wave-length λ, and let CN, or AM parallel to CN, represent the ray direction after leaving the prism. Then MN will be a wave front for the light of wave-length λ. The lens, L, of the telescope, causes the light to converge and there arises a diffraction pattern (not shown) in the focal plane of this lens. The first minimum in this pattern will occur in some direction for which the corresponding diffracted wave front is represented as ME, the angle which ME makes with MN being determined by the fact that the distance NE shall be equal to λ. Further, a wave-length $\lambda' = \lambda + \Delta\lambda$ will be just resolved with reference to λ if an actual wave front for λ' falls along ME, for then the central maximum of the pattern for λ' will fall on the first minimum of the pattern for λ. The ray direction for the light of wave-length λ' is represented by N_0C' in the prism, and by $C'N'$, or AM' parallel to $C'N'$, outside of the prism.

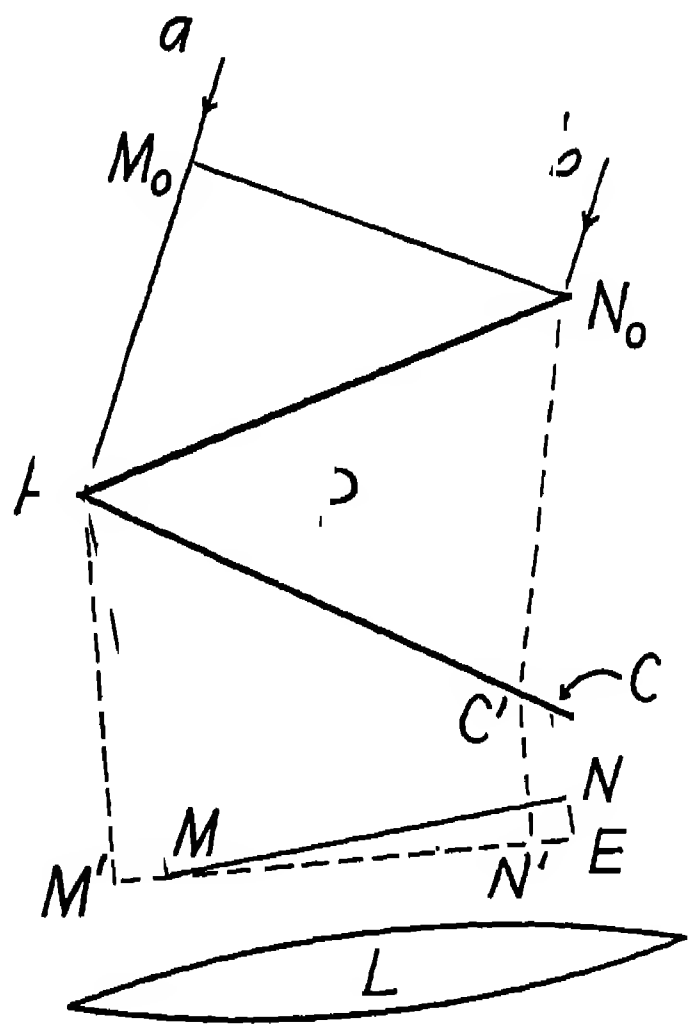


FIG. 165

Let us now consider the optical lengths of the various paths involved: For the sake of simplicity suppose that the prism is ground to a theoretically perfect edge. We may then consider the paths M_0AM and M_0AM' to lie entirely in air. Moreover, it follows from principles deduced in the previous section on the law of extreme path that AM and AM' are equal, to within a small quantity of lower order of magnitude than MM'. Now MM' is of the same order as NE, which is equal to one wave-length. Hence we conclude that the paths M_0AM and M_0AM' are equal to within a quantity which is small compared to the wave-length of light, and we may accordingly consider these paths as equal. Further, denote the index of refraction of the prism for the wave-length λ by μ, and the index for $\lambda' = \lambda + \Delta\lambda$ by $\mu' = \mu + \Delta\mu$. (Since the index of refraction decreases as the wave-length increases, the value of $\Delta\mu$ will be numerically negative—a fact which must be borne in mind.) The paths M_0AM and $\mu N_0C + CN$, for the wave-length λ, are equal, and the paths M_0AM' and $(\mu + \Delta\mu)N_0C' + C'N'$, for the wave-length λ', are equal. Moreover, $M_0AM = M_0AM'$. Therefore, the four paths in question are all optically equal. Equating the two paths which lie near the base of the prism:

$$\mu N_0C + CN = (\mu + \Delta\mu)N_0C' + C'N' .$$

The right-hand side of this equation is the path from N_0 to C' and thence to N' for the wave-length λ'. Let us invoke the theorem that in the vicinity of any actual path the length of path is stationary and substitute for this path a neighboring, sensibly equal path. The foregoing equation then becomes:

$$\mu N_0C+CN=(\mu+\Delta\mu)N_0C+CE\ ,$$

or, breaking up CE into two parts:

$$\mu N_0C+CN=(\mu+\Delta\mu)N_0C+CN+NE\ . \tag{5}$$

Each of the two terms on the left-hand side of this equation may be canceled with an identical term on the right side. After this cancellation we have zero on the left side of the equation, and on the right we have two terms of which one contains the negative term $\Delta\mu$. That is:

$$0=\Delta\mu\cdot N_0C+NE\ .$$

Now N_0C is practically equal to the thickness of the prism at its base, which we shall denote by t, and NE equals λ, hence:

$$\lambda=-t\Delta\mu\ ,$$

or, dividing by $\Delta\lambda$:

$$\frac{\lambda}{\Delta\lambda}=-t\frac{\Delta\mu}{\Delta\lambda}\ . \tag{6}$$

This equation gives the theoretical resolving power of the prism spectroscope which we see is proportional to the thickness of the prism at its base and is independent of the angle of the prism or its altitude. (If the beam is too narrow to fill the prism, there will be a corresponding reduction in the effective value of t.) The resolving power is also proportional to $\Delta\mu/\Delta\lambda$, which is the "dispersive power" of the glass in the wave-length region λ in question. This may be ascertained by looking up, in physical tables, the index of refraction of the kind of glass of which the prism is made, at various wave-lengths. Figure 166 represents the index, μ, for Jena flint glass, plotted as ordinate, against the wave-length, λ, plotted as abscissa. This is a portion of a "dispersion curve."

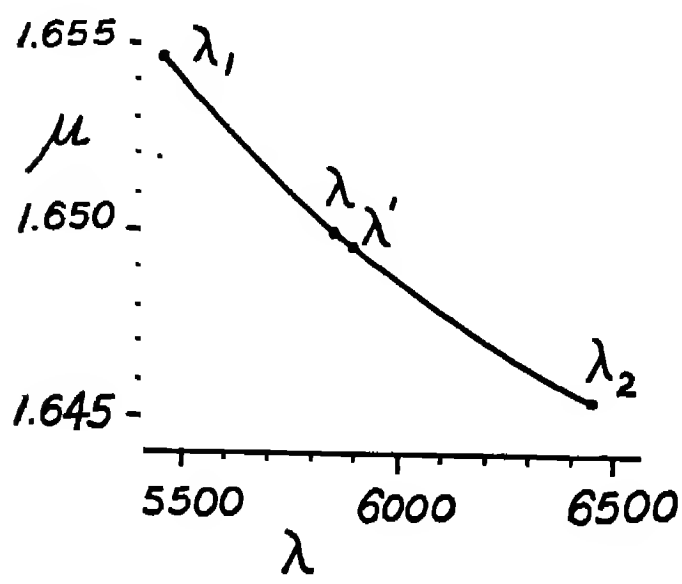


FIG. 166

We shall now take as an illustration the condition for resolution of the sodium lines, $\lambda = 5890$ Å and $\lambda' = 5896$ Å. The wave-lengths λ and λ' are represented in the figure in exaggerated separation: really, $\Delta\lambda = 6$ Å only. The corresponding value of $\Delta\mu$ cannot be satisfactorily ascertained from the dispersion curve. In fact separate values of the index at 5890 Å and 5896 Å will not be found in tables (though Fig. 166 may falsely give the impression that they would be found). However, the dispersive power is the *slope* of the dispersion curve—to be taken in the region where the lines which we wish to resolve are located. This slope can be approximately obtained, either graphically or by choosing two wave-lengths, λ_1 and λ_2, which straddle the lines which we wish to resolve, and then looking up the corresponding indices μ_1 and μ_2. We have:

$$\frac{\mu_2 - \mu_1}{\lambda_2 - \lambda_1} = \text{Slope of dispersion curve} = \frac{\Delta\mu}{\Delta\lambda}.$$

Taking numerical values of wave-lengths and indices for Jena flint glass from Kaye and Laby's *Physical and Chemical Constants*, page 72:

$$\frac{1.6453 - 1.6546}{(6438 - 5461) \times 10^{-8}} = \frac{-.0093}{977 \times 10^{-8}} = -952 \text{ cm}^{-1} = \frac{\Delta\mu}{\Delta\lambda}.$$

Substituting in equation (6), $\lambda = 5890$ Å, $\Delta\lambda = 6$ Å, and $\lambda\mu/\Delta\lambda = -952$ cm^{-1}, we have:

$$\frac{5890}{6} = 952\,t,$$

whence $t = 1.03$ cm. That is, a prism of flint glass about 1 cm thick at the base will just resolve the sodium lines.

The resolving power of the prism spectroscope was first discussed by Lord Rayleigh.*

6. Resolving Power of Microscope.—References and historical remarks will be found at the end of the section.

In illuminating an object under a microscope, a beam from an extended source, such as a frosted incandescent bulb, is commonly reflected from a mirror under the stage of the instrument and directed upward. Let the beam from a certain point of this extended source, after reflection by the mirror, be represented by the slightly divergent beam *ab*, Figure 167; this beam passes through the lenses C_1 and C_2, forming the "substage condenser," and is then focused upon an object point O lying on the axis of the microscope. The condenser is an important accessory. Other points of the source contribute other beams, such as *cd*, and consequently a patch of light

* *Op. cit.*, **1**, 425; **2**, 412; **3**, 105.

is formed in the plane $O'O$. Light, let us say from O, passes through the objective consisting in general of a *train* of lenses, represented in our diagram as L_1, L_2, and L_3. (In a high-power objective the lens L_1 is very small, about a millimeter in diameter.) An image of O, that is, a diffraction pattern p, is formed in the plane PP', and this image is viewed through the eyepiece e, which for simplicity is represented as a single lens. The first minimum in the pattern occurs at P'. An object point O', lying at such a distance from O that the central maximum due to O' falls at P' will be just resolved with reference to O—provided that O' may properly be regarded as an *independent* source. With a telescope two neighboring points may properly be regarded as independent sources. With a microscope there is some question about this, for reasons which will be explained later. However, for the present, we shall discuss the problem in hand as though O and O' were independent sources.

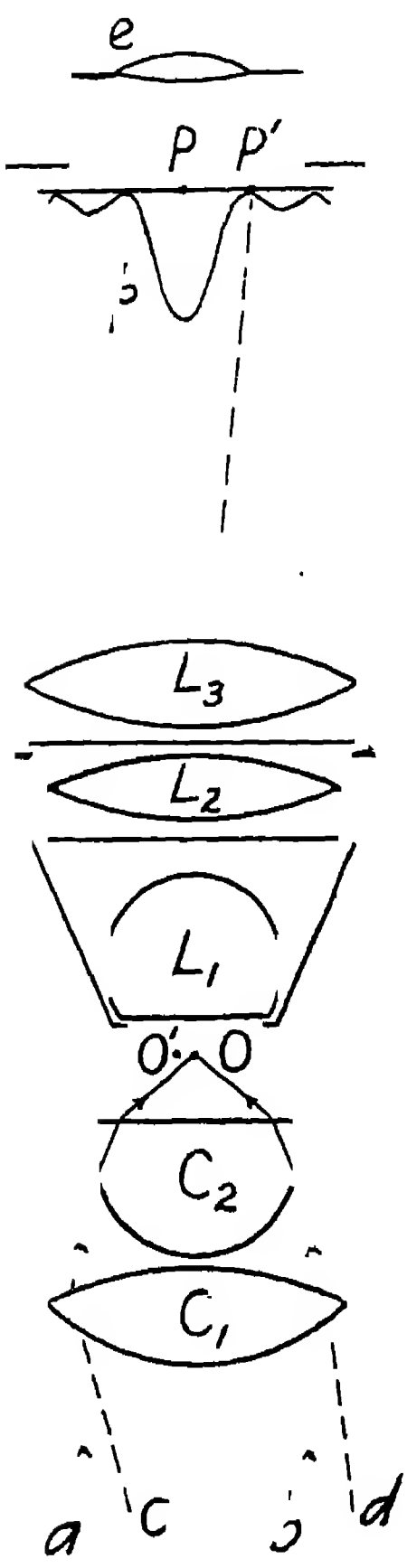


FIG. 167

The resolving power of the microscope is defined as the reciprocal of the least distance which can be resolved. Ordinarily, however, the least distance itself is given, in place of the reciprocal.

The beam incident upon the objective is one of wide angle. The condenser plays its important rôle in forming this beam. The object under examination is not always in air. It may be mounted, for example, in Canada balsam between a lower glass, the "slide" proper, and a thin upper "cover glass." This fact in itself does not require to be taken account of in our theory when the medium between the cover glass and the first lens of the objective, L_1, is air, that is, a medium having an index of refraction $\mu=1$. But with high-power objectives, the space between the cover glass and the lower surface of the lens L_1 is filled, commonly with cedar oil which has an index μ practically equal to that of the balsam and the cover glass ($\mu=1.52$ for sodium light). The hemispherical or overhemispherical lens L_1 also has practically the same index, and the object may then be considered as imbedded in a sphere of index μ. Such an objective is called an "oil-immersion objective." The fact that a refracting sphere has pairs of "aplanatic points" is made use of in the design of the objective. However, this does not directly concern us. What is important for us is that if the theory to be developed is to have general applicability, it is necessary to suppose that the object

may be imbedded in a medium of index μ differing from unity and that all of the space between the object and the objective, and including the first lens of the objective, may have the same index μ. When the beam is to have a very wide angle, a drop of oil is also placed on the upper surface of the condenser lens C_2, and the slide is placed in contact with this drop. On the other hand, the theory will hold also for a "dry objective," a non-immersion objective; in this case the index is simply to be taken as unity.

Referring now to Figure 168, let MHN represent an actual wave front, from O, which is about to enter the objective, or in the case of an oil-im-

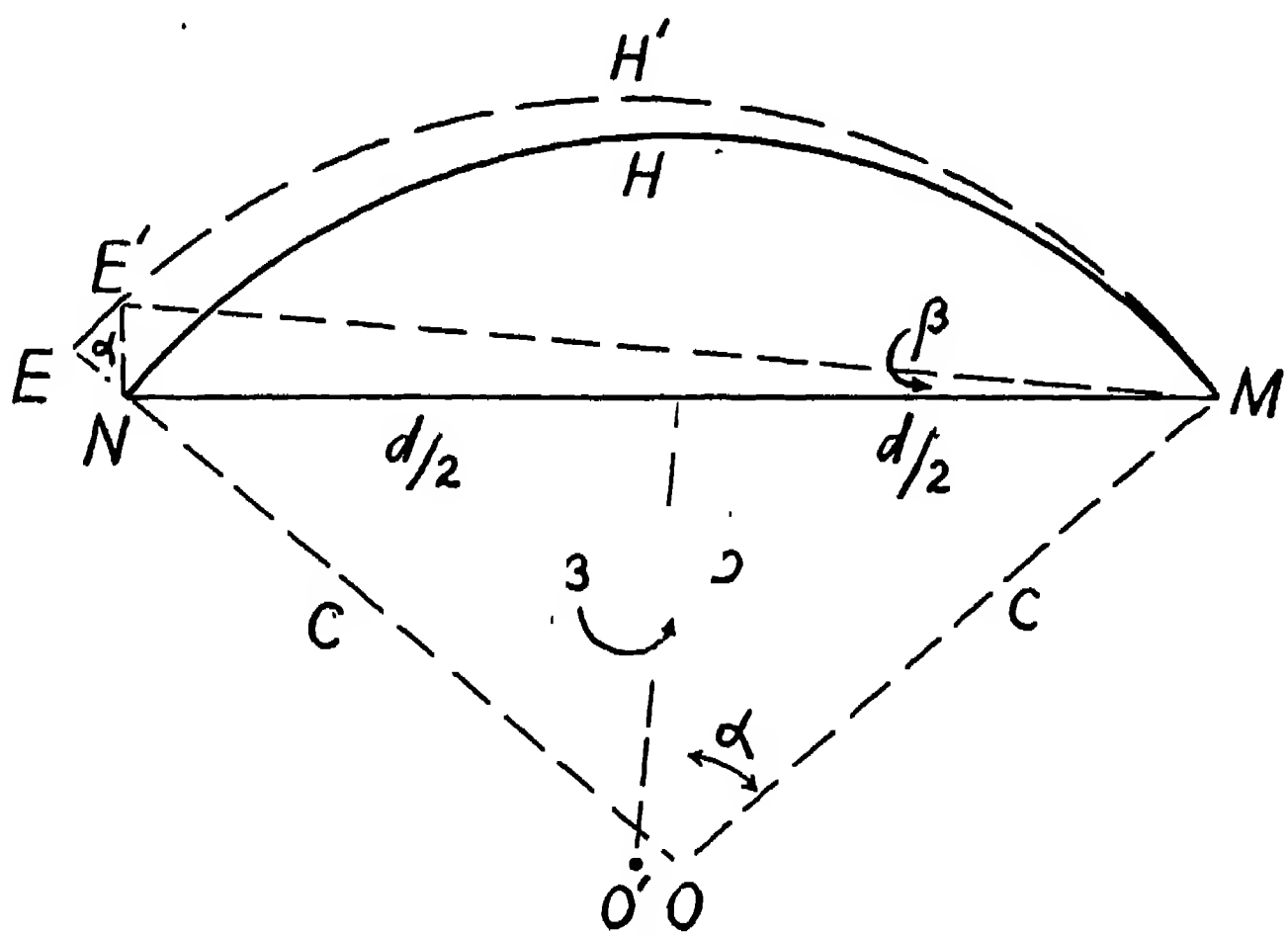


FIG. 168

mersion objective, this may alternatively be after entering the first lens. Let $MH'E$ be the diffracted wave front which corresponds to the first minimum in the pattern. Since the aperture is circular, the distance NE must be equal to $1.22\lambda_0$, where λ_0 is the wave length *in the medium* whatever this may be, of index μ. Let O' be a point just resolved with reference to O; then $MH'E$ must be an actual wave front for O' and, denoting the distance MN by d:

$$\frac{NE'}{d}=\beta=\frac{OO'}{b}. \tag{7}$$

With the telescope the angle OO'/b, the angle subtended by the object points at the objective, is of prime importance. With the microscope we are interested instead in the absolute value of the distance OO'. We shall now find this distance, expressing it in terms *not* involving the distances b or d which are now of only secondary interest.

Denote the semiangle of the beam by α, as indicated. By equation (7):

$$OO' = \frac{b\,NE'}{d} = \frac{b}{d}\left(\frac{NE}{\cos\alpha}\right) = \frac{c\cos\alpha}{d}\left(\frac{NE}{\cos\alpha}\right) = \frac{NE}{2\,\dfrac{d/2}{c}}. \tag{8}$$

Now $NE = 1.22\lambda_0 = 1.22\lambda/\mu$, where λ is the wave-length of the light in air or vacuum. Moreover, $(d/2)/c = \sin\alpha$, hence:

$$OO' = \frac{1.22\lambda}{2\mu\sin\alpha}, \tag{9}$$

where OO' is the least distance which can be resolved.

(This equation is a more general statement of the criterion for resolution than eq. [2] previously found for the telescope; the present equation reduces to eq. [2] when $\mu = 1$ and the angle α is so small that the sine may be considered equal to the tangent.)

The product $\mu \sin\alpha$ is known as the numerical aperture (N.A.). This is an important constant of a microscope objective. The value of μ is commonly 1.52 as already mentioned and when oil immersion is used, the maximum value which α may attain is about 67°. Now $\sin 67° = .92$. Thus, when oil is used the numerical aperture may attain values as high as $1.52 \times .92 = 1.4$. Inserting this value in equation (9), we obtain a least distance of resolution less than $\lambda/2$—less than half a wave-length of light! Moreover, the somewhat arbitrary theoretical factor 1.22 is really too high (see footnote to "Resolving Power of Telescope"). This factor is sometimes taken as low as 1. Also, the value N.A. = 1.4 is not an absolute maximum. By using a medium of high index such as monobromonaphthalene for immersion, and glass of correspondingly high index, a value N.A. = 1.65 can be attained. Inserting in equation (9) the value 1 in place of 1.22 and $\mu \sin\alpha = 1.65$, we obtain $OO' = .3\lambda$. That is, we may hope to resolve two object points separated by only about $.3\lambda$ provided the object points behave as though they were independently luminous.

Two object points separated by a distance as small as $\lambda/2$ or less cannot be regarded as entirely independent sources, because the image of each point of the primary source formed in the plane of the object will spread over a distance equal to the distance between the points. That this must be true may be shown by applying to the substage condenser considerations similar to those above applied to the objective. Imperfection of focus of the condenser will also cause spread of the disturbance in the plane of the object. Thus some points of the source will illuminate both object points simultaneously, but, on the other hand, other points of the source will contribute

to one object-point illumination which the other object point does not share. That is, there will be somewhat of a tendency for the vibrations from the two object points to be in the same phase and to undergo phase changes simultaneously, but there will be no rigorous phase relationship.

Let us next consider the extreme case when the two disturbances, from the two object points, are always in exactly the same phase.

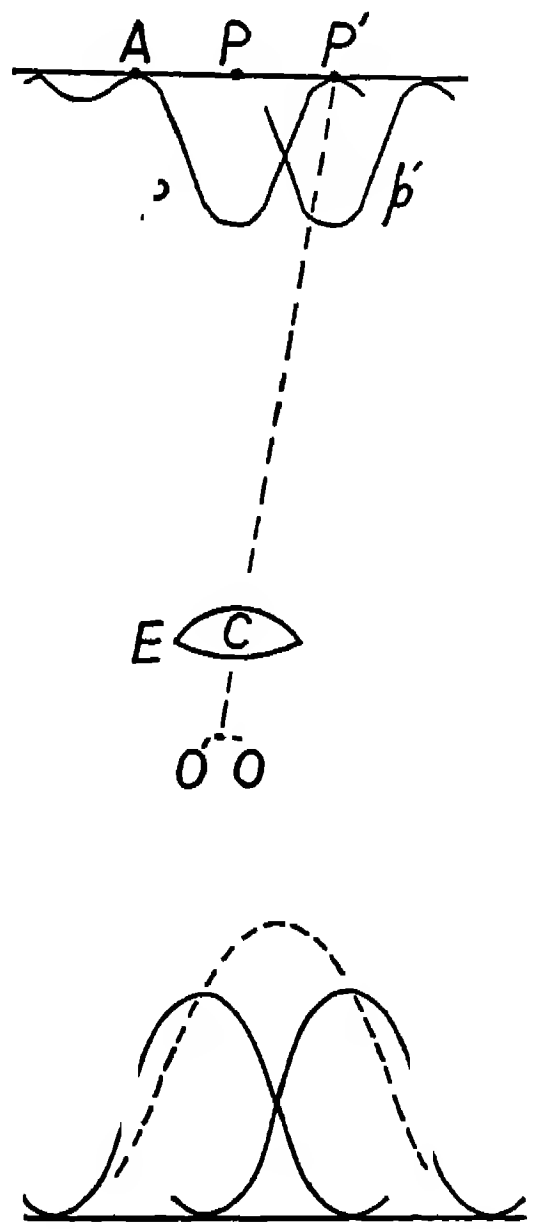


FIG. 169

Referring to the upper part of Figure 169, the image of O will be p. The phase within the interval AP', in the image p, will be everywhere practically the same. This may be seen as follows: The optical distance OCP' differs inappreciably from OCP (as explained in a general way in the concluding paragraphs of sec. 4 and shown in detail in the present footnote).* Also, the optical distance OEP' is greater than OCP', and, accordingly, there is a retardation of phase of disturbances coming by way of the left-hand portion of the lens or lens system. There is a corresponding advance of phase of disturbances coming by way of the right-hand portion of the lens. However, the resultant phase does not change on this account, see, e.g., the vibration curves of chapter 6, Fig. 158, and Appendix C, Figure 265. The resultant phase is the same as the phase of the disturbance coming by way of the center of the lens—within the central maximum—between A and P'. (However, in the first lateral maximum, to the left of A or to the right of P', the phase will be reversed. This becomes evident upon considering the vibration curve to curl up, pass through its first closure, and open again beyond.)

Further, the phase at various points within the central maximum of the image p' is also everywhere the same, and is moreover the same as for p. This follows by steps analogous to those already explained. The phases being throughout, all, always the same, we must add *amplitudes;* the present image curves are amplitude curves. In previous cases the image curves were intensity curves, and we added these, this being the proper procedure

* If $OO' = \lambda/2$ and the objective magnifies 100 times, then $PP' = 50\lambda$, or, taking λ as 5×10^{-5} cm, we have $PP' = 2.5\times10^{-3}$ cm. Take $CP = 10$ cm, then applying the sagittal relation, $CP' - CP = (PP')^2/2CP = (2.5\times10^{-3})^2/20 = 3.12\times10^{-7}$ cm. I.e., the path increment is small compared to the wave-length, and therefore the corresponding phase change will be negligibly small.

when there is no constant phase relation between the two images. The sources, when independent, each undergo phase changes independently many more times per second than the eye can follow, and then the averaged-out resultant is the sum of the intensities. But now we must, on the other hand, add the amplitude curves. These are shown to a larger scale, and erect, in the lower part of the figure. It is to be noted that the height of each curve halfway between its central maximum and first minimum is *more* than *half* of the height of the maximum. (This holds true both for a circular and for a rectangular aperture.) Hence when the two amplitude curves are added we obtain a curve, indicated by the broken line, which has a maximum at its center (instead of a slight dip, as when adding intensities). Thus, the images are evidently too close for resolution when the object points are in phase. However, if the separation were increased until the two minima coincided, the images would evidently be resolved, with a little to spare. To have the minima coincide requires double the previous separation, and we have then:

$$OO' = \frac{1.22\lambda}{\mu \sin \alpha}, \tag{10}$$

where OO' is (nearly) the least distance which we can resolve when the object points emit vibrations in the same phase. For given numerical values of the constants equation (10) leads to double the values of equation (9).

Actual practice shows that when a beam of the proper wide angle is furnished by the condenser, least distances of resolution smaller than those given by equation (10) can be obtained. The distances conform more nearly to equation (9). It is possible to resolve points separated by a distance of only $\lambda/2$, and even points separated by a somewhat smaller distance.*

The over-all magnification of the instrument must be sufficiently great so that the final beams can be comfortably resolved by the eye. The eye cannot resolve beams subtending angles less than about one and one-half minutes of arc.†

The so-called "ultra-violet microscope" gains its advantage by resorting to illumination by light of short wave-length. This instrument has lenses entirely of quartz. Because of the impossibility of achromatizing when only one substance is used for the lenses, the ultra-violet light used for illumination must be roughly monochromatic.

* The one-celled alga *Amphipleura pellucida* reveals striated markings at an interval only 2.5×10^{-5} cm which can still be clearly resolved. For quantitative tests of the limit of resolution with objectives of various numerical apertures see B. K. Johnson, *Jour. Roy. Microsc. Soc.*, **48**, 144, 1928. The method used, however, inherently precludes the use of the highest numerical apertures.

† For further information regarding desirable magnifications see Martin and Johnson, *Practical Microscopy* (Blackie & Son, 1931), p. 36.

Our discussion has thus far had to do with the possibility of resolving two points close together and, by inference, with the possibility of seeing detail in an object. This possibility has nothing to do with how small a particle may be and its presence still be subject to detection under the microscope. This latter question depends principally upon whether the particle will scatter a sufficient amount of light, when the illumination is introduced horizontally; and this in turn depends upon how intense it is possible to make the primary beam. Here the "ultra-microscope" comes in. The ultra-microscope must not be confused with the ultra-violet microscope. It is not an instrument unto itself, but is a high-power microscope of the usual type used in conjunction with a system of horizontal illumination which yields an intense beam having a very small thickness, measured vertically. No direct light enters the microscope. Each particle (suspended in a transparent medium) acts as a center of scattering and gives rise to an image, a diffraction pattern, which is seen against a dark background. The pattern is practically the same as that which would be formed by a geometrical point. It gives, by its size or form, no clue to the size or form of the object; there is no question of *resemblance* to the object. However, an idea of the relative size of the scattering centers may be obtained by comparing the relative amounts of light which they scatter, that is, by comparing the *intensities* of the diffraction disks to which they give rise. It is possible, by means of the ultra microscope, to detect the presence of, for example, colloidal particles much smaller than $\lambda/2$, and it is possible, by counting the number in the field at one time, to determine the concentration of a colloidal suspension.

A method of measuring the diameters of microscopic particles is discussed in section 9.

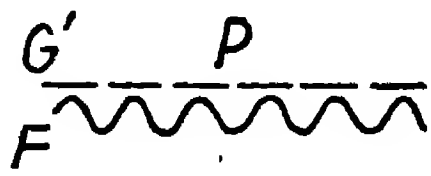
The development of the microscope was given great impetus by the long fruitful labor of E. Abbe. In the year 1873 he stated conclusions which he had reached in regard to the possibilities of resolution.* In the following year H. von Helmholtz† independently discussed the resolving power of the microscope. The subject was subsequently discussed notably by Lord Rayleigh.‡ The treatment above given bears more resemblance to that of Rayleigh than to that of Abbe or von Helmholtz. Abbe's point of view for the case of vibrations in the same phase deserves special attention and forms the subject of our next section.

* *Gesammelte Abhandlungen*, **1**, 45, also *Die Lehre v.d. Bildentstehung im Mikroskop* (a posthumous book compiled and edited by O. Lummer and F. Reiche; Vieweg: Braunschweig, 1910).

† *Pogg. Ann.: Jubelband*, p. 557, 1874.

‡ *Op. cit.*, **4**, 235, and **5**, 118.

7. Abbe's Theory of Microscopic Vision.—Following Abbe, let us suppose that the object viewed under the microscope is a diffraction grating, G (Fig. 170), composed of alternate transparent and opaque strips. The objective of the microscope, represented by the lens L, forms an image of the grating, at G'. If the grating is illuminated by a parallel beam of monochromatic light incident perpendicularly from below, as indicated by the beam i, the disturbances leaving the transparent strips, the object points, will all be in the same phase.



Abbe answers the question of whether or not the lines of the grating constituting the object will be resolved, upon the basis of whether or not the beams which are diffracted by the grating pass through the microscope or fail to pass.

The grating will form the beam of zero order, *0*, the beams of first order, *1* and *1'*, the beams of second order, *2* and *2'*, etc. Each of these beams, being parallel, will come to a focus in the upper principal focal plane of the objective, that is, at I_0, I_1, etc. The points I_0, I_1, etc., are principal diffracted images in monochromatic light of the original source which is to be thought of as a point, or a line parallel to the grating lines, lying at infinity. Or, if the source lies at a finite point, a collimating lens must be used.

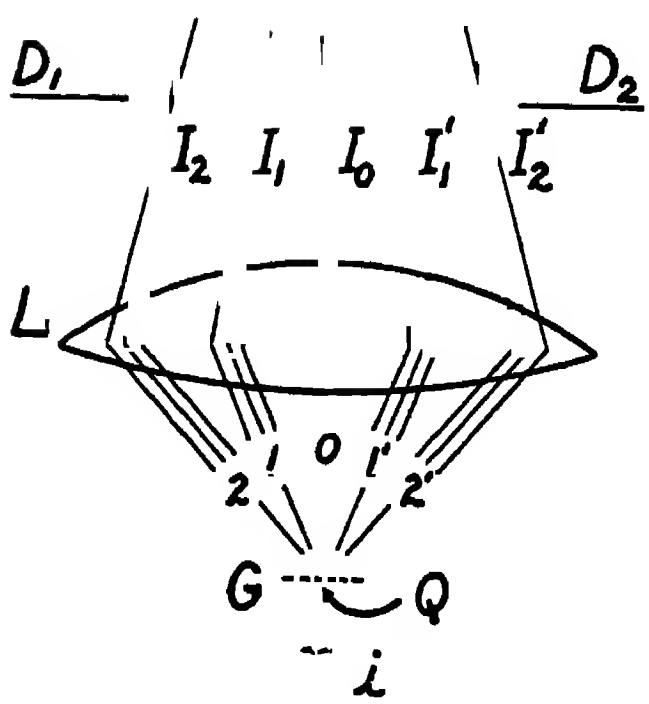


FIG. 170

The central transparent space, Q, of the grating lies presumably on the axis of the microscope, and hence the image of this space, P, will also lie on the axis. The optical distances from Q, respectively by way of I_0, I_1, etc., to P are all equal. The basic concept of Abbe's theory is that we may regard the image, G', of the grating as an interference pattern, formed by interference of the disturbances coming from I_0, I_1, etc.

Let D_1 and D_2 be diaphragms which may be moved inward to cut off such beams as we may choose. Suppose first that both diaphragms are moved in so that only the central beam, I_0, is allowed to pass. Obviously this beam will spread out uniformly in any plane beyond, in particular in the plane $G'P$. There is then no chance for interference. The lines of the

grating, the points of the object, will not be resolved. If, however, D_1 is withdrawn until the beam I_1 is also allowed to pass, the images I_0 and I_1 will, together, form a set of Young's interference fringes, as represented by the curve F. (In this curve the maxima are represented as downward, i.e., there is a maximum at P.) It can be shown (see reference given in footnote)* that the spacing of these fringes is independent of the particular wave-length of the light used and, what is important, that the spacing is equal to the spacing which obtains in the image. The fringe system F may thus be regarded as a poorly resolved image of the grating. Further, the images I_0 and I_1' acting in conjunction would form the same fringe system as I_0 and I_1. The diaphragms D_1 and D_2 determine, by their positions, the effective angular aperture of the objective. In the microscope as actually used the aperture is symmetrical with reference to the axis. When the aperture is of such small angle that the beams I_1 and I_1' are both excluded, there is no resolution. When these beams are both admitted, we have resolution. We will then have three interfering beams, and the resultant interference or diffraction pattern will be formed somewhat in analogy with the formation of the pattern by a three-line diffraction grating, the images I_0, I_1, and I_1', playing the part of the lines of the grating. Care must be used, however, in applying this analogy. The absence of a focusing lens between the analogues of the grating lines, namely, I_0, I_1, and I_1', and the field plane $G'P$ does not upset the analogy, because the microscope objective takes care of the fact that the optical paths QI_0P, QI_1P, and $QI_1'P$ are all equal, as they should be. However, the image I_0 will in general have a distinctly greater intensity than I_1 and I_1', and this fact will cause the resultant pattern to differ from that for a three-line grating. The dominance of I_0 may suppress the faint intermediate maximum which occurs between the principal maxima in the pattern due to a three-line grating. Thus when all three images I_0, I_1, and I_1' are active, the resulting pattern may again resemble our present curve F.

If α denote the semiangle of the beam included by any objective, and θ_1 denote the angle of diffraction of the beam *1*, or *1'*, of order $m=1$, then, for resolution, α must be at least as great as θ_1, and the limiting condition for resolution becomes $\alpha=\theta_1$. The grating law, under the given circumstances, is $\lambda_0=\lambda/\mu=a \sin \theta_1$, where λ_0 is the wave-length in whatever medium the grating may be immersed, λ is the corresponding wave-length in air or vacuum, μ is the index of the medium, and a the grating space. When $\alpha=\theta_1$ the microscope will just resolve the distance a, hence we have for the least distance of resolution $a=\lambda/\mu \sin \theta_1=\lambda/\mu \sin \alpha$. If the aperture of the system is circular we should, perhaps mainly for the sake of consistency, introduce

* Mueller-Pouillet, *Lehrb. d. Phys.* (9th ed.), **2**, Part I, 705.

the factor 1.22. We then have, for the least distance of resolution, when the disturbances all leave the grating in the same phase:

$$a = \frac{1.22\lambda}{\mu \sin \alpha}, \tag{11}$$

which is identical with equation (10), which also applies for two object points emitting disturbances in the same phase.

When the second- and higher-order beams are admitted, the image becomes progressively more perfect and will be sensibly perfect when, and only when, all of the beams having appreciable intensity are included in the beam admitted by the objective.

If we block off all of the images of *odd* order, I_1, I_1', etc., there will remain the images, of even order, I_0, I_2, I_2', etc. The remaining interfering sources I_0, I_2, I_2', etc., will then have *double* the previous separation of the interfering sources and will as a consequence produce a system of fringes having *half* of the previous spacing. That is, if all of the existing beams of even order are admitted, and only these, we shall have a sharp image, and, what *appears* to be a perfect image, but this will be a *false* image of the object. We shall then have the true image of a grating having a grating space $a' = a/2$, which, as we know, would have the beams *2* and *2'* for its first-order beams, the beams *4* and *4'* (not shown) for its second-order beams, etc. In other words, the image which we obtain is a true image of that grating, that object, which would give rise to all of the beams and only the beams which are admitted.

Let us now consider the original beam to be incident at an angle, let us say from the right, in such a direction that the undeviated beam will continue as beam *1*. Suppose the diaphragms D_1 and D_2 are pushed in until they just let I_1 and I_1' pass. The beam I_1 is now the zero order, I_0 is the first order for a grating of spacing a, and I_1' is the second order. The object, the grating, will evidently be resolved, with something to spare. Further, a grating twice as fine, having a grating space $a' = a/2$, would again have I_1 for its beam of zero order, but then the beam I_0 would be absent, and I_1' would be the first-order beam. This grating would be just resolved. Thus if the light is incident at such an angle that it just enters the objective, we can resolve a grating twice as fine as before, and hence we have for the least distance of resolution:

$$a' = \frac{1.22\lambda}{2\mu \sin \alpha}. \tag{12}$$

In the case under consideration, the disturbances leaving the various lines of the grating differ in phase among themselves but are still "coherent."

Equation (12) is identical with equation (9) for independently luminous object points.

Abbe checked his theory experimentally and numerous others have likewise done so. A. B. Porter* devised an interesting and simple demonstration bearing upon the theory. Referring to Figure 171, S represents a point source of light which for simplicity of discussion we may suppose to be monochromatic. The light from S passes first through a convex lens L, then through a coarse crossed diffraction grating, G, a wire gauze, and is finally received in the plane $JKMN$, which is focally conjugate to the plane in which the source lies. The crossed grating forms a central image at 0; a horizontal row of spectrum dots of various orders, 1, 2, $1'$, $2'$, etc.; a vertical row of spectrum dots, a, b, a', b', etc.; and diagonal rows along the directions JN and KM (see chap. 6, sec. 13, on "Crossed Gratings"). Since the grating is coarse, the spectrum dots all lie close together, and it is easy to arrange matters so that they all fall within an area which can be taken in at once by the pupil of the eye (or alternatively by a small telescope). When the eye is used unaided, it is placed centrally in the plane $JKMN$ or, let us say, just behind this plane. When the pupil takes in all of the diffracted beams of appreciable intensity, one sees the gauze as it really is. Now suppose a diaphragm with a slot in it is placed in the plane $KJMN$. When the slot is horizontal and admits only the horizontal row of images 0, 1, 2, $1'$, $2'$, etc., one apparently sees that grating which would give rise to these images and to these only, namely, a coarse grating of vertical wires. The horizontal wires of the gauze are not seen. When the slot is turned vertically, only the horizontal wires are seen. If the slot is turned so as to take in one of the diagonal rows of images only, one thinks one sees wires along the diagonal perpendicular to the slot. Finally, if an aperture is made in the form of a cross and placed so as to take in both diagonal rows of images, one apparently sees a gauze with wires running in both diagonal directions, a gauze which is not present at all.

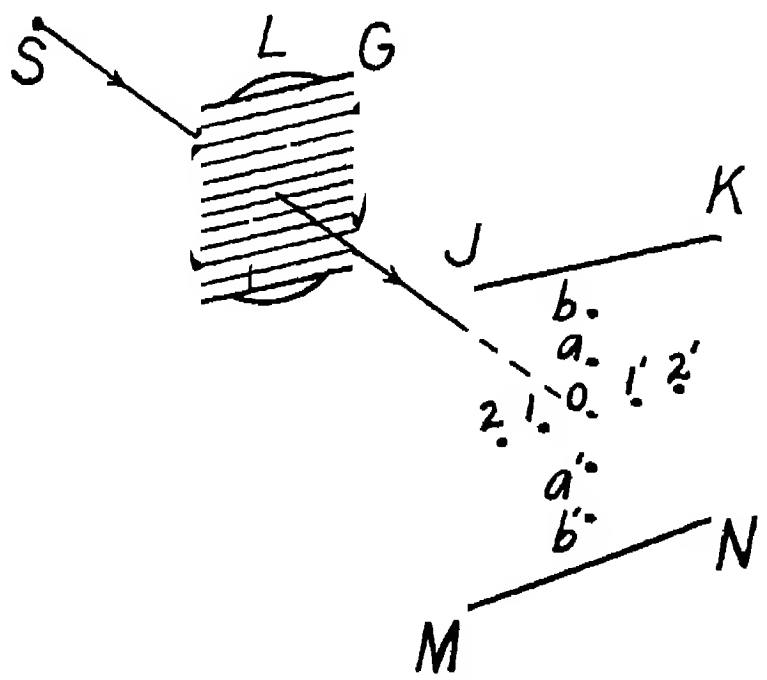


FIG. 171

There has been much controversy over whether Abbe's theory is necessary, or sufficient, or both, when the incident beam fills a cone of wide angle, as it does, and should, in the ordinary use of the microscope. There are

* *Phil. Mag.*, **11**, 154, 1906.

many supporters of the so-called "equivalence theory" who hold that an object illuminated by a cone of wide angle is fully equivalent to a self-luminous object.

Be it pointed out that the mere fact that we exclude no beams is not sufficient to assure resolution. A grating of grating space less than λ, when illuminated perpendicularly, forms no diffracted beams—only an undeviated beam. Yet if the microscope takes in this beam alone the grating will evidently not be resolved.

A list of references on Abbe's theory is given at the end of the book compiled by Lummer and Reiche (1910; see footnote at end of our sec. 6). A recent discussion is by H. Moore (*Jour. Roy. Microsc. Soc.*, **48**, 133, 1928), who refers especially to M. Berek (*Marburger Sitzungsberichte*, December, 1926).

8. Measurement of Star Diameters.—The method of measurement to be described was suggested in basic principle by H. L. Fizeau* as early as 1868 and was tried in 1874 by E. Stephan,† who correctly concluded that the angular diameters of the fixed stars were too small to measure—by the method in the state to which he had developed it. Important improvements were suggested in 1890 by A. A. Michelson,‡ but not until 1920 did he bring the method to fruition for determining the diameters of any of the fixed stars. Even so, only giant stars subtend an angle which is sufficiently large to be measurable. The method, now known as "Michelson's method," may also be used in simpler form to determine the angle subtended by double stars, or, by way of a laboratory exercise, to determine, for example, the separation of two narrow slits close together.

Referring to Figure 172, the objective lens, L, of a telescope is covered by a cap having two rectangular apertures M and N of which the separation may be varied.

The image of the source is not, as such, a matter of interest.

To begin with, suppose a far-distant linear source of monochromatic light of wave-length λ lies in the direction O.

The apertures M and N each constitute a wide slit and, in the principal focal plane F of the objective, there will be formed a Fraunhofer diffraction pattern for a double slit. (The formation of this type of pattern was ex-

* *Compt. rend.*, **66**, 934, 1868.

† *Ibid.*, **76**, 1008, 1873, and **78**, 1008, 1874.

‡ *Phil. Mag.*, **30**, 1, 1890; *Amer. Jour. of Sci.*, **39**, 115, 1920; *Astrophys. Jour.*, **51**, 257, 1920 and **53**, 249, 1921 (the last with F. G. Pease); *Science*, **57**, 703, 1923; also *Light-Waves and Their Uses* (1903) and *Studies in Optics* (1928), both published by the University of Chicago Press.

A comprehensive account is given by F. G. Pease, *Ergebnisse d. Ex. Naturw.*, **10**, 84, 1931.

plained in detail in chap. 5, sec. 5.) Referring now to the upper part of our present Figure 173, suppose each aperture acting alone yields the image, the diffraction pattern, represented by the broken curve AB. When the two apertures are effective simultaneously, the pattern AB fills in with narrow fringes, P_0, P_1, etc., as indicated by the solid curve. These fringes arise from interference between the disturbances from the two apertures, the resultant disturbance from each aperture being considered as a unit. The maxima P_0, P_1, etc., occur where the difference of the paths, from the center of each aperture to the field point in question, is 0, λ, 2λ, etc. Denoting the separation of the apertures, between centers, by s and the focal length of the objective by F, the angle subtended by adjacent maxima, for example, P_0 and P_1, at the objective, is $\alpha = P_0P_1/F = \lambda/s$. Since F is constant, the angle α which obtains at any given separation s of the apertures may be directly indicated in the diagram (see top of figure).

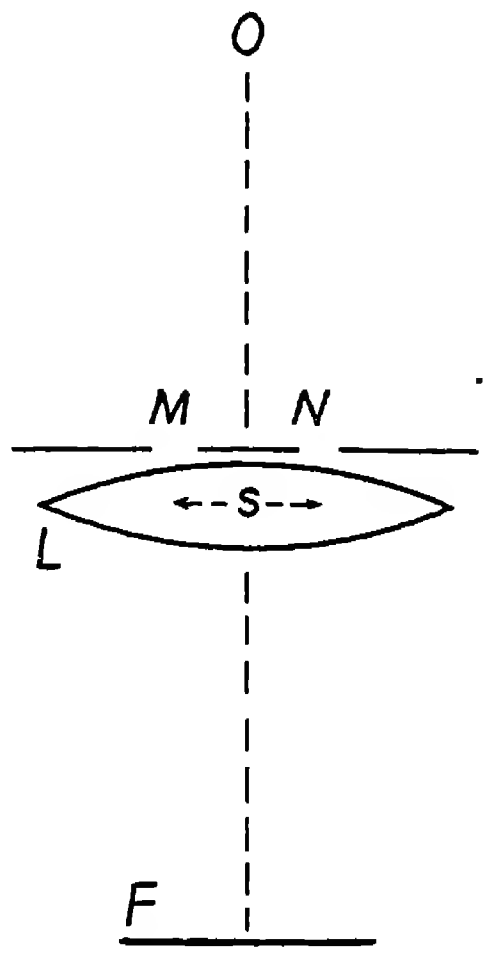


FIG. 172

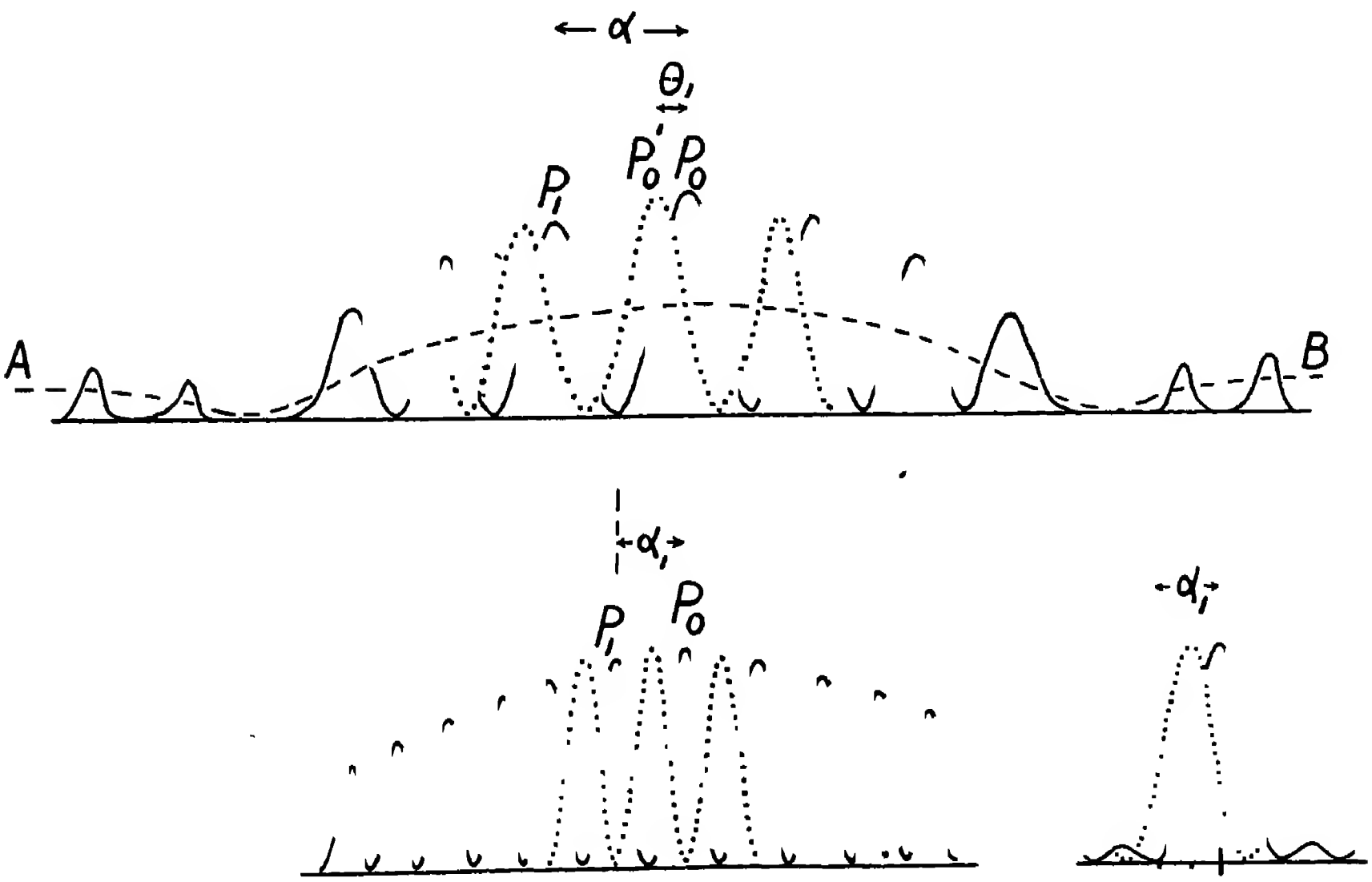


FIG. 173

Suppose now that there is a second linear source to the right of the first one and close to this and parallel to it, the two sources subtending a small

angle θ_1 at the objective. The second source will give rise to a set of fringes similar to the first set, only displaced to the left, as represented by the dotted curve P_0' in the upper part of the figure. The distance P_0P_0' by which the second curve is displaced is such that $P_0P_0'/F=\theta_1$. Under the supposed conditions, the separation of the two sets of fringes is considerably smaller than the distance between adjacent maxima of one set, and in this case the only effect will be somewhat of a decrease in the contrast which obtains between the maxima and the minima of the summational curve, a decrease in the so-called "visibility" of the resultant fringes, but these will still be easily observable.

Now let the separation of the apertures M and N be gradually increased. As this separation increases, the angle $\alpha=\lambda/s$ between adjacent maxima of each fringe system decreases. The central maxima, P_0 and P'_0, always maintain their positions (see lower central portion of figure); new fringes crowd in from both sides of the field. When the separation of the apertures attains a value s_1 such that the angle subtended by adjacent maxima of each set of fringes, for example, P_0 and P_1, is $\alpha_1=\lambda/s_1=2\theta_1$, the maxima of the first set of fringes, solid curve, will fall on the minima of the second set, dotted curve, and vice versa, and the resultant fringes formed by the two sets disappear. Thus, if s_1 is the separation of the apertures at which the fringes disappear, the angle subtended by the two sources is:

$$\theta_1=\frac{1}{2}\frac{\lambda}{s_1}. \tag{13}$$

By way of comparison let us consider the performance of a telescope having a single wide rectangular aperture of width s_1 equal to the separation of the two previously considered apertures at which the fringes disappear. The telescope with this single wide aperture would form images of the two sources, as represented in the inset in the lower right-hand corner of Figure 173. In each image the angle from the central maximum to the first *minimum* would be $\alpha_1=\lambda/s_1=2\theta_1$, where θ_1 is as before the angle subtended by the sources. This angle α_1 is the same as the angle between adjacent *maxima* in the other case. But in the present case we fall short of resolution, because the angle from central maximum to first minimum exceeds the angle of separation of the two images. It exceeds this by a factor of 2. We are short of resolution by this factor. Thus, when it is desired to measure the angle subtended by two sources close together, the use of the two apertures over the objective yields a gain of a factor of 2 as compared to direct observation. Moreover, disappearance of the fringes can be more precisely ascertained than resolution.

Returning to the case of the two apertures: After the fringes have dis-

appeared, separating the apertures further will cause them to reappear. Let us suppose that the diameter of the telescope is great enough to admit of indefinite further separation of the apertures. The fringes attain a second maximum of visibility when they get so narrow that the first maximum on the left, of the first set, falls on the central maximum of the second set. This occurs when the separation of the apertures is $2s_1$. Subsequently a second disappearance occurs at the separation $3s_1$. Several successive reappearances and disappearances can be obtained. The separation at the first disappearance being s_1, the disappearances occur at separations s_1, $3s_1$, etc., and the maxima of visibility at separations $2s_1$, $4s_1$, etc.

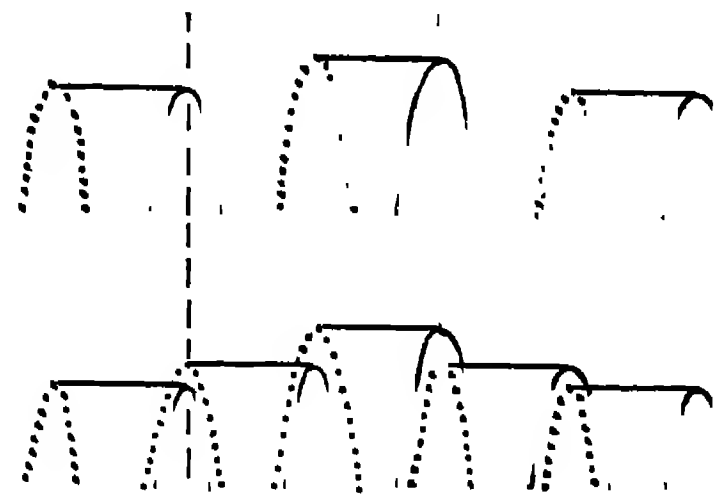


FIG 174

The theory as so far considered applies to double linear sources and also to double stars.

Let us now suppose that we have a source consisting of a vertical *band* of monochromatic light. Referring to the upper portion of Figure 174, the band of light will cause a spread of each maximum over an "interval," as indicated by the horizontal lines joining the maxima represented in the figure. Denote the angle subtended by the band by θ. A double linear source of angle $\theta_1 = \theta$ would cause the fringes to disappear when the apertures are separated by a distance s_1 given by the equation $\theta_1 = \theta = \lambda/2s_1$. At this separation of the apertures the situation, when viewing the luminous band, is as represented in the upper portion of the figure. The distance between maxima of consecutive order is twice as great as the spread of each maximum. The fringes will remain, but with lowered visibility. However, when the separation of the apertures attains a value $s = 2s_1$ the fringes will have contracted until the distance between maxima of consecutive order is equal to the spread of each maximum (see lower portion of figure). In this circumstance, overlapping causes the fringes to disappear. We now have $\theta = \lambda/(2s_1) = \lambda/s$, where s is the separation of the apertures at the first disappearance of the fringes due to a luminous band, and θ is the unknown angle subtended by the band. Or:

$$\theta = \frac{\lambda}{s}. \tag{14}$$

When the separation of the apertures is carried beyond the value s, the fringes reappear and subsequently attain a lowered maximum of visibility

when the separation is $3s/2$. This is followed by a second disappearance at the separation $2s$, etc.

Let us now consider as source a circular disk which is uniformly luminous over its surface, of which the diameter subtends an angle θ' at the objective of the telescope. Vertical strips taken on the disk have decreasing height as we approach the limb, hence the angle θ' must be greater than the angle which a luminous band subtends when the fringes disappear in both cases at the same separation s of the apertures. That is, $\theta' > \theta = \lambda/s$. Theory shows that the familiar factor 1.22 enters here. That is:

$$\theta' = 1.22\,\frac{\lambda}{s}, \tag{15}$$

where s is the separation of the apertures at the first disappearance of the fringes, when the source is a uniformly luminous circular disk.

A hitherto unmentioned, very important feature of the method, introduced by Michelson, will now be explained. Referring to Figure 175, a system of mirrors M_1, M_2, M_3, M_4 may be mounted close in front of the telescope objective. The effective separation of the apertures now becomes the distance s between the centers of the mirrors M_1 and M_2, and the attainable effective separation of the apertures is thus greatly increased. The theory outlined above still applies, without alteration. However, the distance between the mirrors M_3 and M_4 plays an important secondary rôle. Let AN be an actual wave front and AE a diffracted wave front, the distance NE being, for example, λ. The angle between these two wave fronts will then be λ/M_1M_2. After reflection from the mirrors M_3 and M_4, these wave fronts become respectively $A'N'$ and $A'E'$. The distance $N'E' = NE = \lambda$. The corresponding angle is now λ/M_3M_4. That is, all angles are magnified in the ratio M_1M_2/M_3M_4, and consequently the fringe spacing undergoes a much-needed magnification, for the spacing would otherwise be excessively small when the distance M_1M_2 is large. When the arrangement of four mirrors is used, a large telescope is not an inherent necessity. However, for the measurement of star diameters a large telescope

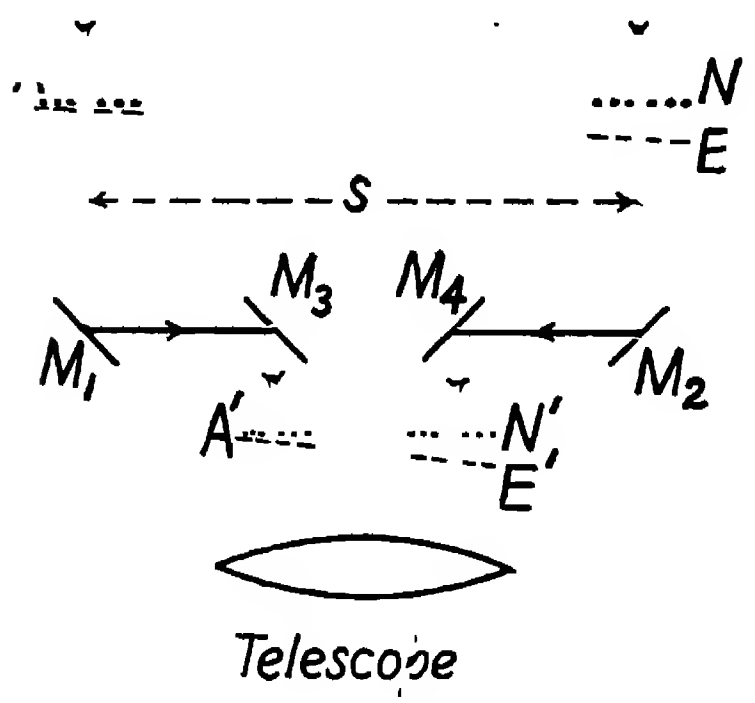


FIG. 175

must nevertheless be used in order to secure the requisite mechanical strength and stability in the mounting of the entire system.

There is another important advantage of Michelson's method. When a large telescope is used in the ordinary manner, with full aperture, the ever varying optical inhomogeneity of the atmosphere often causes so-called "boiling" in the field of view. On the other hand, when two comparatively small apertures are used, differential atmospheric variations over the two apertures ordinarily cause only a slow minute drifting of the entire fringe system which is readily followed by the eye.

Measurements of the diameters of a number of giant stars have been carried out by Michelson, ably assisted by F. G. Pease, using the 100-inch telescope of the Mount Wilson Observatory. By way of example it was found that the red giant star Betelgeuse (α Orionis) causes the fringes to disappear at a separation of the mirrors M_1 and M_2 of about 300 cm. Taking λ as 5750 Å, the angular diameter of Betelgeuse becomes $\alpha' = 1.22[5750\times10^{-8}/300] = 2.3\times10^{-7}$ radians (=.047 seconds of arc), presupposing uniform luminosity over the surface. The distance of this star from the solar system, independently determined by the method of parallax, is about 1.7×10^{20} cm. Hence the linear diameter, still supposing uniform luminosity, is $(2.3\times10^{-7})(1.7\times10^{20}) = 3.9\times10^{13}$ cm, or about 240,000,000 miles, or about 300 times the diameter of the sun. However, Betelgeuse is a star of exceedingly low density and in stars of this type the luminosity no doubt decreases toward the limb of the disk. Hence the actual diameter is supposedly even greater than the figure given. Arcturus, Mira, and Antares are still bigger giants than Betelgeuse.

In the measurement of star diameters, elaborate accessory apparatus is necessary for the purpose of making and maintaining adjustments.

The vast majority of stars have an angular diameter so small that it cannot be measured even by Michelson's powerful and beautiful method.

For further information regarding the measurement of star diameters see especially the article by F. G. Pease referred to in footnote on page 216.

9. Measurement of Diameters of Microscopic Particles.—The basic principles of the method described in the previous section may be applied for measuring the diameters of small particles under a microscope. A suggestion to this effect was made first by Michelson in 1890 and subsequently by others, but actual trial was not made until 1926, when experiments were carried out by U. Gerhardt.*

The two requisite rectangular apertures of variable separation are most conveniently placed, not between the object and the objective, but above the objective, that is, between this and the image plane. A given separation

* *Zeit. für. Phys.*, **35**, 697, 1926; **44**, 397, 1927.

of the apertures determines a corresponding value of the effective numerical aperture of the optical system, at this separation. However, the effective value of the numerical aperture, as determined by the separation of the two rectangular apertures, can never exceed the actual numerical aperture of the objective. And herein lies a serious limitation of the method when applied to the microscope. With the telescope, when mirrors are mounted on a girder, as previously described, the aperture of the objective no longer sets a limit. But the numerical aperture of the microscope objective always sets a limit. However, it is possible to measure diameters down to about 1.5×10^{-5} cm, which is somewhat lower than the smallest distance of resolution which can be attained by direct observation, and it is moreover possible to attain a fair degree of accuracy in the measurement of such small diameters.

10. Luminosity of the Diffracting Edge (Observation).—Let us suppose that light from a point source falls upon a screen bounded by a single vertical straight edge and consider what an observer stationed beyond the screen will see when looking toward the edge. Let the screen, as seen by the observer, extend to the left. Taking up our position within the shadow, let us suppose that we are viewing the edge through a small telescope of a type which does not invert. A telescope permits of more definite focusing than the unaided eye, and this is an advantage, for reasons which will appear. The telescope should be diaphragmed down. An aperture of about 5 mm will be found suitable when observation is being made at a distance of 1 or 2m.*

Upon focusing the telescope upon the diffracting edge, the edge appears as a luminous line of definite length. When we are within the shadow as supposed, the appearance is as represented at the left of Figure 176. Here DD' represents the straight edge of the screen. This is luminous over a length AB. The luminosity is most intense at the center of the line segment in question and decreases toward the top and bottom of this. To indicate this fact the line segment is drawn heavier in the middle than at the ends. However, this device of representation is not to be taken as meaning that the luminous segment itself is broader in the middle than at the ends. The segment appears simply as a geometrical line having no observable width, or at least no width which is not attributable to the fact that the telescope does not form a perfect image. It is to be remembered that a telescope cannot form a perfect image because diffraction always sets a limit to the perfection of any image.

* If the telescope does not permit of focusing at such close range, a spectacle lens may be placed over the objective. A "bird glass," such as can be purchased from the large mail-order houses for about one dollar, serves well as an erecting telescope. There is, however, no real objection to the use of an inverting telescope, but it will be found less convenient.

If the diffracting screen were entirely removed, we would see the point source of light in the field of the telescope, but the image would be decidedly out of focus. We would have a so-called "extra-focal" image which would be a circular patch, represented in the figure by the circle O drawn in broken line. The height of the luminous line which we see when the screen is in

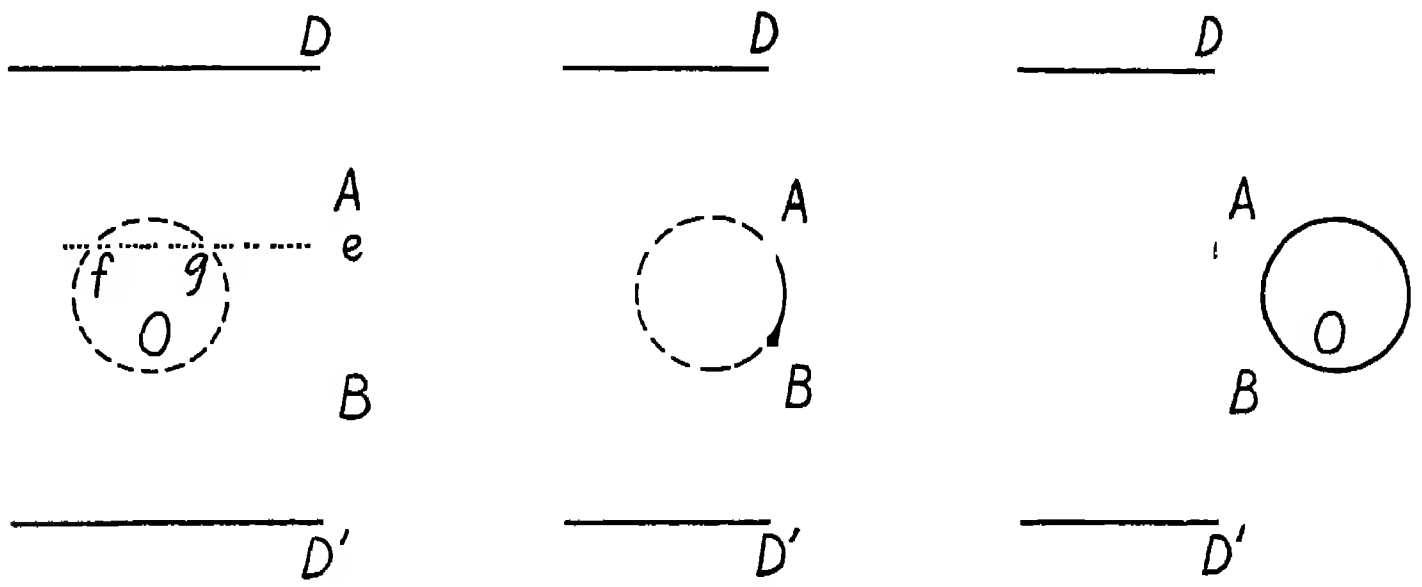


FIG. 176

place is equal to the height, the diameter, of this circle, and, further, the intensity at each point of the line segment AB is proportional to the width which the patch has along the normal to the edge erected at the point in question. For example, at the point e the intensity is measured by the width fg of the patch. The diameter of the circular patch itself is determined by the aperture of the diaphragm which is supposedly being used before the objective of the telescope.

As we move, from a given position, farther into the shadow, the luminous line AB becomes rapidly fainter, disappearing under ordinary conditions of observation at an angle of diffraction of some 5°. Toward the end of the range, the color of the line becomes distinctly red in hue, because the longer wave-lengths are inflected farther than the shorter ones. It is possible by resorting to a different method of illumination to observe the luminosity of the edge at even quite large angles of inflection. We shall concern ourselves with observations at large angles in a later section, but not at present.

Let us now suppose that we move, from a given initial position within the shadow, toward the edge of the shadow. As we proceed, the luminous line becomes more intense and soon the circular patch begins to protrude from behind the straight edge, as shown in the middle portion of Figure 176. The luminous line can no longer be distinguished where it falls on the protruding portion of the extra-focal image, but we can still observe luminosity along the upper and lower portions of the line segment AB, as illustrated. There is now, however, some glare which makes observation more difficult.

When we have proceeded so far that the extra-focal image is entirely clear of the straight edge, as shown on the right-hand side of the figure, we shall again observe the line of luminosity along the full length AB.

Very instructive and beautiful observations may be made by setting up the diffracting system represented in Figure 177. A copper cent, C, is fastened to the end of a wire, $GHJK$, of somewhat over 1 mm diameter, bent to form a roughly but not accurately circular arc as indicated. The wire is then fastened to a diffracting screen having a straight edge. The center of curvature of the arc into which the wire is bent, in so far as there is a center, should lie on the screen, let us say at the position O'. The telescope should be easily movable both vertically and horizontally so that the extra-focal image of the source may be made to occupy any position in the field at will. A mechanism which permits both vertical and horizontal motion and at the same time keeps the telescope always directed at the diffracting system is convenient. When the extra-focal image occupies the position represented by the broken circle O, the image itself being obscured from view, we may expect to see the diffracting edges luminous along AB, and at E, F, G, H, J, and K, that is, wherever a perpendicular to the edge passes through the extra-focal image of the source. However, at points far removed from O, especially at F, the amount of light may be too feeble to observe. If now, by moving the telescope, the circular patch is moved to O', the wire will appear luminous along the full length of both its inner and its outer edges in a very striking manner.

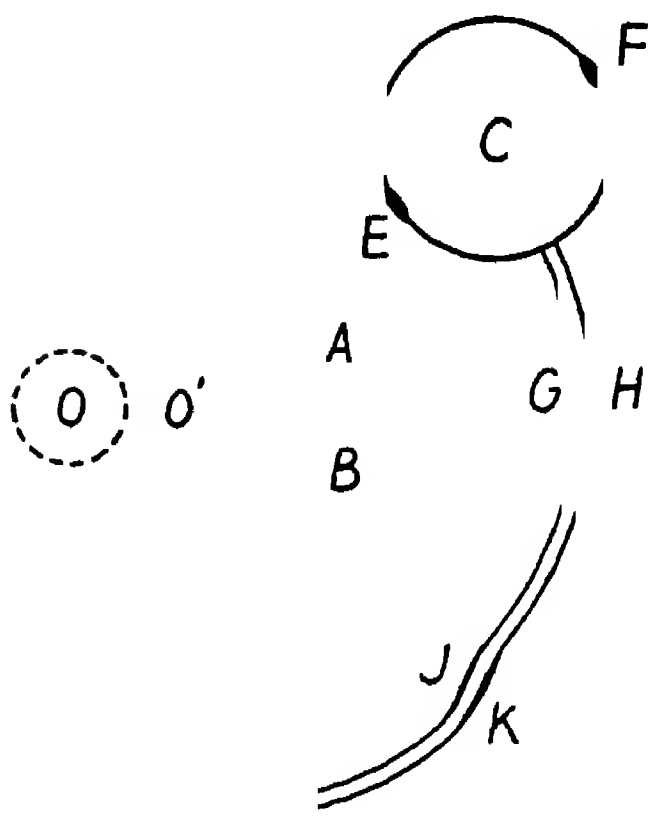


FIG. 177

If the telescope is moved to such a position that the extra-focal image of the source falls behind C, the entire circumference of the coin will appear luminous. In addition, the straight edge will be luminous at a point opposite C. The extra-focal image does not need to be centered accurately behind C. The entire circumference of the coin remains luminous while the patch is so located that it includes the center of the coin within itself.

Stating the facts of observation in a new manner: The luminous portions of the edge occur wherever the optical path is either a maximum or a minimum or is constant. Fermat's principle, or law of extreme path—or, better, the law of stationary path—applies to diffraction phenomena. The paths in question are to be chosen from among all paths taken from the point source by way of various points of the diffracting edge or edges to any one point at a time of the aperture of the telescope, or, when we are observing with the

unaided eye, to any one point at a time of the pupil of the eye. Wherever the path is stationary, there the edge appears luminous.

11. Luminosity of the Diffracting Edge (Theory).—Let us first consider in detail the case in which the point source of light is located at infinity. The incident wave fronts are then plane. Referring to Figure 178, let W be an incident wave front which is just passing the diffracting edge D, and let us suppose that we are looking toward the edge from some point P_s within the shadow. The fact that the edge D appears luminous means that the disturbances arising from the wave front W must build up a wave front having a cross-section which is a circle described about D as center. Let the arc AB represent such a wave front. Since the amount of light which we receive on passing into the shadow diminishes rapidly, it follows that the amplitude which obtains on the wave front AB must diminish rapidly in passing from A to B. This circumstance is indicated in the figure by tapering off the width of the arc as drawn.

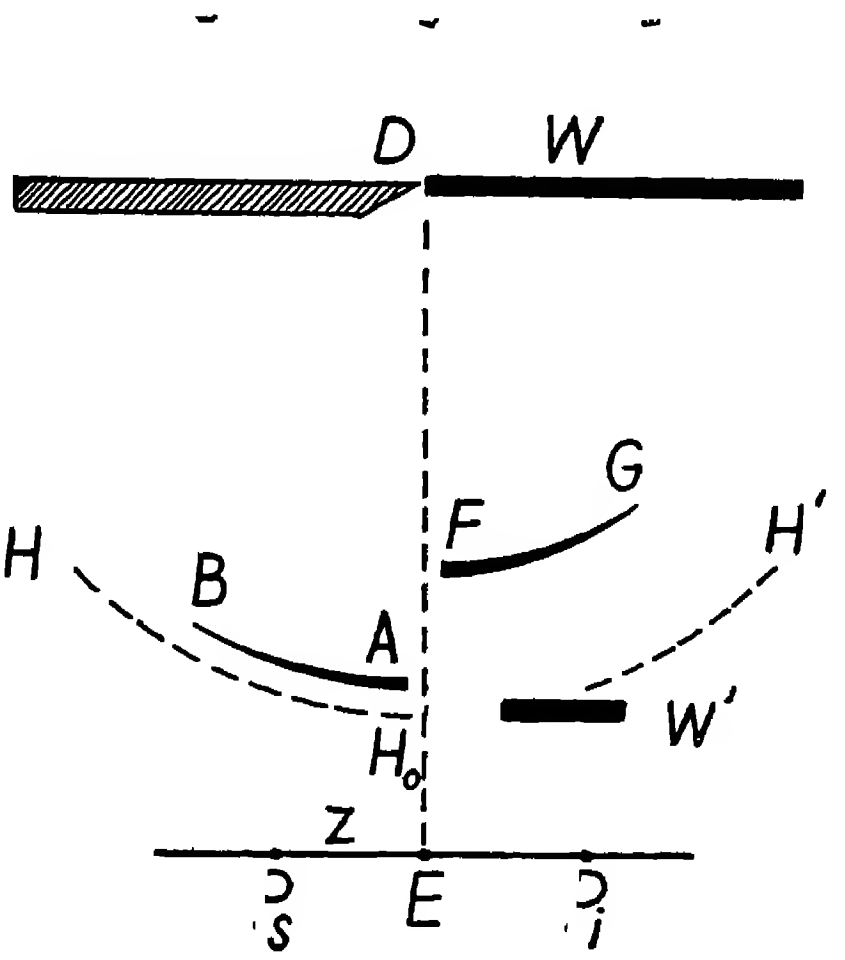


FIG. 178

A Huygens wavelet originating from the diffracting edge D, or really from the point of the incident wave front which is in contact with D, would also have a cross-section which is a circle described about D as center. Let the broken arc HH_0H' represent such a wavelet.

It will be proved later that the inflected wave AB is circular about D as center and that it has suffered a retardation of 1/8 of a period with reference to the wavelet HH_0H'. Accordingly, the distance from the arc AB to the arc HH_0 is to be regarded as being equal to $\lambda/8$.

When we take a position at some point P_i in the illuminated region and focus on the diffracting edge, we obtain exactly the same extra-focal image which we would obtain if the diffracting screen were absent, and in addition we see the edge D as luminous. In view of this circumstance we must conclude that the telescope analyzes the total incident light into two parts, one part being a wave front W', which is the wave front which would be formed if the diffracting screen were absent, and the other part being a wave front

FG having a horizontal cross-section which is circular with D as center. This wave front, like AB, must diminish rapidly in amplitude with increasing angle of diffraction because the amount of light which we receive from the edge D in moving to increasing angles of diffraction diminishes rapidly. The deflected wave FG will be shown later to have suffered a retardation of phase of five-eighths of a period with reference to the wavelet HH_0H'. Accordingly, the distance between the arcs FG and H_0H' is to be regarded as being equal to $5\lambda/8$.

The problem which lies before us is to explain why the incident wave front W gives rise to the wave fronts AB, W', and FG. We should explain this if possible upon the basis of the same principles which we have applied in explaining diffraction in general. These principles, as previously worked out, lead, in the present instance, to the Cornu spiral applied as a vibration curve. Accordingly, we shall seek to explain the formation of the wave fronts in question by having recourse to the Cornu spiral.

It will be recalled that distances, v, measured from the origin of the spiral along the arc, are linearly proportional to distances, z, measured in the field plane from the edge of the geometrical shadow to the field point in question. (The constant relating v and z is given in eq. [3] of chap. 4.)

A disturbance which reaches the edge of the geometrical shadow E from the diffracting edge D is represented on the Cornu spiral by an infinitesimal element of arc taken at the origin of co-ordinates, that is, by an elementary vector pointing horizontally to the right. Further, a disturbance reaching a point P_s within the shadow, from the diffracting edge D, is represented by an element of the spiral taken in the upper branch of this at the proper distance from the origin. This disturbance will be retarded in phase with reference to the disturbance reaching E by an amount $\psi = \pi v^2/2$, where v is the distance from the origin along the upper branch of the spiral which corresponds to going into the shadow a distance $z = EP_s$. In Figure 179 the broken curve H_0H represents the retardation of phase ψ which is in question. Here ψ is plotted as ordinate, against v plotted as abscissa. This curve is simply a plot of $\psi = \pi v^2/2$ against v, that is, the curve is a parabola. The foregoing refers to the disturbance originating from the diffracting edge.

The disturbance which the whole effective portion of the incident wave front causes at the edge of the geometrical shadow is represented on the Cornu spiral by the vector drawn from the origin of the spiral to the upper asymptotic point, or, in the notation of the spiral, by the vector OJ. This vector represents the resultant disturbance in both amplitude and phase.*

* The Fresnel theory leads to an inconsistency in regard to phase, as was previously pointed out in chap. 2, sec. 12. This inconsistency takes the form of ascribing the wrong phase to the disturbance which the effective portion of the wave front causes at a given field point. However, in taking phase differences in the manner in which we shall take them, the error in question is eliminated.

The disturbance which the incident wave front causes at the above-considered point within the shadow is represented in amplitude and phase by the vector drawn from the point lying at the foregoing distance v along the upper branch of the spiral to the upper asymptotic point J. This vector rapidly shortens as we pass farther into the shadow, and this shortening accounts for the rapid diminution in amplitude of the inflected wave with increasing angle of diffraction.

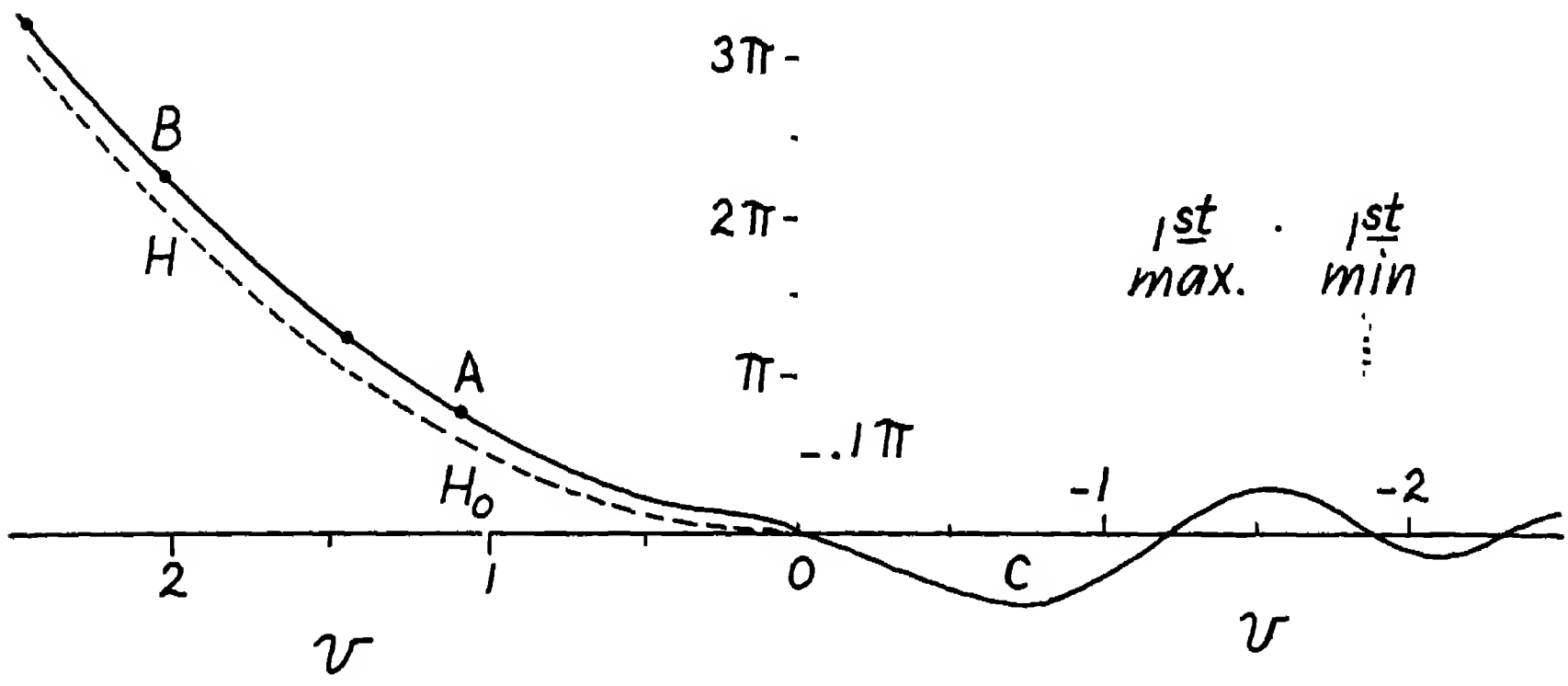


FIG. 179

The resultant disturbance reaching a given point within the shadow will be retarded in phase with reference to the disturbance reaching the edge of the geometrical shadow by an amount which is measured by the angle of rotation which the closing vector to the upper asymptotic point of the spiral sweeps out as we pass from the edge of the geometrical shadow to the given point within the shadow. In Figure 179 the solid curve AB represents the loss in phase now in question, plotted as ordinate, against distance into the geometrical shadow, in terms of v, plotted as abscissa. Denote the loss in phase by ϕ. By inspection of the spiral (Fig. 46, in chap. 3) it will be seen that when v takes the value 1.07 the closing vector to J is directed horizontally to the left. The corresponding value of ϕ is therefore $\phi=\phi_{1.07,\,J}-\phi_{0,\,J}=\pi-\pi/4=3\pi/4$; accordingly, $v=1.07$ and $\phi=3\pi/4$ are the co-ordinates of one point, A, of the solid curve of Figure 179. Similarly, when $v=1.44$, the closing vector to J is directed vertically downward and hence $\phi=3\pi/2-\pi/4=5\pi/4$; hence another point is to be plotted with co-ordinates $v=1.44$ and $\phi=5\pi/4$. In this manner the solid curve may be plotted point by point. Since the Cornu spiral soon convolves in nearly circular arcs, the solid curve takes a course which is, to a high degree of approximation, parallel to the dotted curve—the vertical distance between the two curves is practically the same at various points. This distance has the value $\pi/4$, that is, for example, $AH_0=BH=\pi/4$. Now the dotted curve represents a

disturbance diverging from the diffracting edge as center, and since the two curves differ by a constant amount in phase, the solid curve must also represent a disturbance diverging from this edge as center, and moreover this second disturbance must be retarded by the constant amount in question, namely, $\pi/4$ or $\lambda/8$, with reference to the first one. In other words, the effective portion of the incident wave front gives rise to an inflected wave front which appears to originate from the diffracting edge and is retarded by an amount $\lambda/8$ with reference to a disturbance traveling from the source by way of the edge.

The constancy of the difference $\pi/4$ or $\lambda/8$ holds to a high degree of approximation until we approach very close to the edge of the geometrical shadow. Here the distance between the two waves rapidly varies from $\lambda/8$ to zero, as evidenced by the approach of the solid curve of Figure 179 to the dotted curve upon approaching the origin of co-ordinates. This variation from $\lambda/8$ to zero may be interpreted as an aberration, and from what we have learned regarding permitted aberration we may conclude that the variation from $\lambda/8$ to zero will cause only a very slight deterioration in the perfection of the image of the luminous edge as observed. However, such deterioration as occurs will take the form of a broadening of the image on one side, and this broadening will pass over, without discontinuity in the physical phenomenon, into the protrusion of the extra-focal image of the source from behind the diffracting edge when we move slowly toward the illuminated region.

We shall now consider what happens in the illuminated region. As we pass from the edge of the geometrical shadow into the illuminated region, the phase of the disturbance at first advances with respect to the constant phase which would prevail in the field plane in the absence of the diffracting edge. This circumstance is indicated in Figure 179 by the fact that the curve C, which is a plot of *loss* of phase in this region, at first takes negative values. The advance of phase in question is deducible from the Cornu spiral: As we pass from the origin of the spiral to points in the lower branch, the closing vector to the upper asymptotic point becomes less inclined to the horizontal than the vector OJ, or the vector $J'J$ which represents "normal phase" in the field plane; this lessened inclination represents advance in phase. The first minimum in the phase curve, C, representing maximum gain of phase, occurs when $v = -.6$, approximately, as may be ascertained by inspection of the spiral. This minimum occurs about halfway out to the point in the figure which is marked "1st max.," this being the point where the first maximum in the amplitude or intensity pattern occurs. This maximum occurs when $v = -1.2$, approximately. Similarly, the point designated as "1st min." represents the position of the first minimum in the amplitude

or intensity pattern. For reasons of clarity of representation the right-hand half of Figure 179 has been drawn on the basis of a vertical scale which is several times larger than that used on the left side of the figure. That the phase curve C will take the form shown in our figure is deducible directly from the Cornu spiral; we need not go into further detail.

Referring now to Figure 180, the vector CJ represents the effect at a given field point in the illuminated region. We may consider this vector as the sum of the two vectors CJ' and $J'J$. As we move farther out into the illuminated region, the point C moves farther out along the lower branch of the spiral. At all stages we have a corresponding vector sum, $CJ = CJ' + J'J$, the vector CJ' varying and $J'J$ remaining constant. The vector $J'J$ represents the effect which would obtain if the diffracting screen were absent. But because the diffracting screen is present, the Huygens wavelets which would originate from the portion of the incident wave which is stopped by the screen are absent. The resultant effect of these wavelets is the effect which we would have at the same field point if the actual diffracting screen were replaced by its complementary, that is, by a screen extending from the actual diffracting edge indefinitely to the right, instead of to the left as we have been supposing. This effect is represented by a vector drawn from J' to C, or, by the negative of CJ'; in place of our foregoing sum we may write $CJ = J'J - J'C$. That is to say, the total actual light may be regarded as being equal to normal illumination, $J'J$, less the wave which would be inflected by the complementary screen. And therefore, since the inflected light would appear to come from the diffracting edge, the deflected light will also appear to come from this edge, and the fact that the deflected light is the negative of the inflected, means that the deflected light must be considered to have suffered a retardation of one-half period more than the inflected, or 5/8 period in all, or $5\lambda/8$.

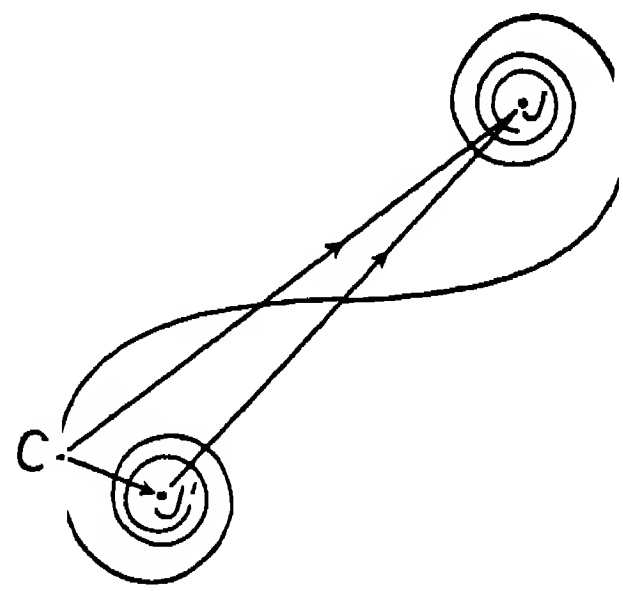


FIG. 180

The physical phenomenon is continuous in crossing the edge of the geometrical shadow. The illumination varies continuously in crossing this edge, and the phase likewise varies continuously. However, we change our point of view abruptly at the edge of the geometrical shadow. Upon approaching this edge from within the shadow, we consider the direct light to be zero until we reach this edge, and there we consider the direct light to change discontinuously to full value. When, in our thought process, we introduce a discontinuity in what we more or less artificially designate as the direct

light, we must, in order to remain in keeping with the physical continuity which actually prevails, introduce a counterbalancing discontinuity in regard to what we more or less artificially designate as the wave front diffracted by the edge. This counterbalancing discontinuity takes the form of considering the deflected wave to be reversed in phase with regard to the inflected wave. There is no physical discontinuity.

It is surprising upon first thought that the eye—or, let us say, a telescope—situated in the illuminated region and directed toward the diffracting edge should analyze the total light which is incident upon the telescope according to the rather artificial scheme which has just been outlined. Our surprise, however, diminishes when we consider that if an actual source of light, properly constituted and vibrating in the proper phase, were inserted at the position of the diffracting edge, and the diffracting screen were then removed, the original point source and this inserted source would together yield exactly the same wave front as before; and the telescope would of course analyze this wave front in such a manner as to reveal that the total light originates from two separate sources.

The external fringes which are formed in the case under consideration may be regarded as arising from interference of the light deflected by the edge with the direct light. This is the manner in which Thomas Young regarded the external fringes as being formed, as recounted in chapter 1. Thus Young's theory of diffraction is found, after all, to contain an important element of truth. However, the deflected wave front does not arise from ordinary reflection by the edge, as Young supposed. The diffracting screen acts merely as a barrier, and the deflected wave is one component of the whole disturbance occurring in the region beyond the edge, the whole disturbance being accounted for by summing up, in accordance with the principle of interference, the effects of the Huygens wavelets from all of that portion of the incident wave front which passes the diffracting screen.

The external fringes soon merge into normal illumination because the amplitude of the deflected wave decreases rapidly with increasing angle of diffraction.

The case in which the point source is situated at a finite distance may be treated in much the same way as when the source lies at infinity—as we have thus far supposed. Let a be the distance from the source to the diffracting edge and b the distance from this edge to a given field point lying on the edge of the geometrical shadow. In the horizontal plane containing the source describe a circle about the source as center with a radius $a+b$. If the diffracting screen were absent, one and the same phase would prevail at all points of this circle. Upon insertion of the diffracting edge the phase at each point of the circle would be altered by an amount depending upon the

position of the point along the circle. Now it can be shown from the Cornu spiral, in exactly the same way as for plane incident waves, that the phase which prevails at each point of the circle within the geometrical shadow is retarded by $\pi/4$, or $\lambda/8$, with reference to the phase of a Huygens wavelet originating from the diffracting edge. Hence the inflected wave again appears to originate from the diffracting edge. And in a fashion also similar to that for plane incident waves it can be shown that there is a deflected wave which is retarded by $5\pi/4$, or $5\lambda/8$, with reference to a Huygens wavelet originating from the edge.

When the incident waves are plane and the direction of incidence is perpendicular to the diffracting edge, the diffracted wave is a cylinder having the edge as axis; hence the radius of curvature of the diffracted wave in the vertical plane equals infinity. When the incident waves are diverging, the diffracted wave has a radius $a+b$ in the vertical plane, and a radius b in the horizontal plane, where a as before denotes the distance from the source to the diffracting edge, and b denotes the distance from the edge to any arbitrary point where we may be considering the diffracted wave front.

When, as always in optical diffraction experiments, we observe at distances from the diffracting screen which are large compared to the wave-length, the rays, the lines of energy flux, appear to radiate from the diffracting edge. We are not justified, however, in projecting these lines backward as straight lines to within a few wave-lengths of the edge. Our theory does not justify us in drawing any conclusions concerning the course which the lines of energy flux take near the edge. There are several objections to drawing such conclusions; the apparent luminosity of the diffracting edge is a consequence of the shortness of the wave-length of light. In an acoustical experiment, if we were observing within three or four wave-lengths of the diffracting edge, the wave would not appear to originate from the diffracting edge with anything like the same high degree of approximation which obtains in the optical experiment which we have been considering.

It has been attempted in the foregoing paragraphs to present as simple an exposition as possible of an important aspect of diffraction theory which has hitherto been available only in highly technical and mathematical treatments. These treatments show that a surface integral taken over the diffracting aperture permits of mathematical transformation into a line integral taken around the edge of the aperture. This line integral has values only where the path from the source by way of the edge to the given field point is either a maximum or a minimum or is constant; and experiment reveals that this mathematical transformation also has physical significance —when the eye or a telescope is placed at the field point in question the diffracting edge appears luminous at the points where the path is stationary.

When the diffracting edge is curved, the inflected wave is retarded by an amount which differs from that for a straight edge, $\lambda/8$. But the wave deflected to the same field point by the complementary diffracting screen is always retarded by an additional amount of $\lambda/2$. This will appear more fully toward the end of the next section, which deals with Babinet's principle.

For references to the more highly technical treatments see footnote.*

Any diffraction pattern, of either the Fresnel class or the Fraunhofer class, which admits of analysis, may be reinterpreted from the viewpoint which has just been outlined. The direct light is to be considered as having the value of normal illumination everywhere outside of the geometrical shadow and as being equal to zero within this. Superposed upon this distribution and added to it in accordance with the principle of interference is the light diffracted by the edge. For example, an irregular obstacle combined with a somewhat larger circular aperture always yields a maximum of illumination on the axis of the aperture. This fact was pointed out at the end of chapter 2 and was there explained on the basis of the appropriate vibration curve. By way of comparison, our present viewpoint calls for an explanation somewhat as follows: The direct light is zero everywhere behind the obstacle, but a point on the axis of the aperture will receive a maximum of diffracted light because light will be deflected to this point from the entire periphery of the circular aperture, and this light will all arrive in the same phase.

The zone plate furnishes another good example. The maxima which occur along the axis may be interpreted as occurring where the light inflected from the inner edge of a given transparent zone is reinforced by the light deflected at the outer edge of the same zone. For example, at the first-order maximum there is a geometrical path difference of $\lambda/2$ to which is to be added an amount of $\lambda/2$ because the deflected light is retarded by this amount with reference to the inflected light. There is thus a total retardation of λ and hence these two disturbances will reinforce each other—and these disturbances will be reinforced by similar pairs of disturbances from the inner and outer edges of all of the other transparent zones.

A diffraction grating composed of equal transparent and opaque strips

* A summarizing discussion and numerous references are given in Geiger and Scheel, *Handb. d. Phys.*, **21**, 906–11. In the same connection see Försterling, *Lehrb. d. Optik* (Leipzig: Hirzel, 1928), p. 254. Among those who contributed to what we may call the revival of Young's theory in modified form are to be mentioned especially. E. Maey, *Ann. d. phys.*, **49**, 69, 1893, and **73**, 16, 1924; also *Zeitschr. f. d. Phys. u. Chem. Unterricht*, **17**, 10, 1904, and **36**, 96, 1923; A. Sommerfeld, *Math. Ann.*, **47**, 317, 1896, and in Riemann-Webers, *Diff. Gl. d. Phys.* (7th ed.), **2**, 468; and A. Rubinowicz, *Ann. d. Phys.*, **53**, 257, 1917, and **81**, 140, 1926.

furnishes a similar example. At the first-order maximum the geometrical path difference between corresponding points of adjacent strips is λ. Hence the disturbances inflected at the near edges of all of the transparent strips will reinforce each other and these disturbances will moreover be exactly reinforced by the disturbances deflected at the far edges of the transparent strips because the latter disturbances suffer a total retardation of λ of which $\lambda/2$ is attributable to geometrical path difference and the remaining $\lambda/2$ is attributable to additional retardation upon deflection. From our present point of view we can see without recourse to vibration diagrams why a grating composed of equal transparent and opaque strips yields more light in the first order than when the strips are of unequal width. When the strips are unequal, the inflected and the deflected disturbances fail to reinforce each other exactly, and hence the maximum has lower intensity.

12. Babinet's Principle.—Two diffraction screens, A and B (Fig. 181), are said to be complementary when the opaque portions of screen A correspond to the apertures of screen B and vice versa. Babinet's principle has to do with the diffraction patterns formed by complementary screens. What the principle tells us depends upon conditions. We shall first take the simplest case, which is also the case which leads to the most striking conclusion.

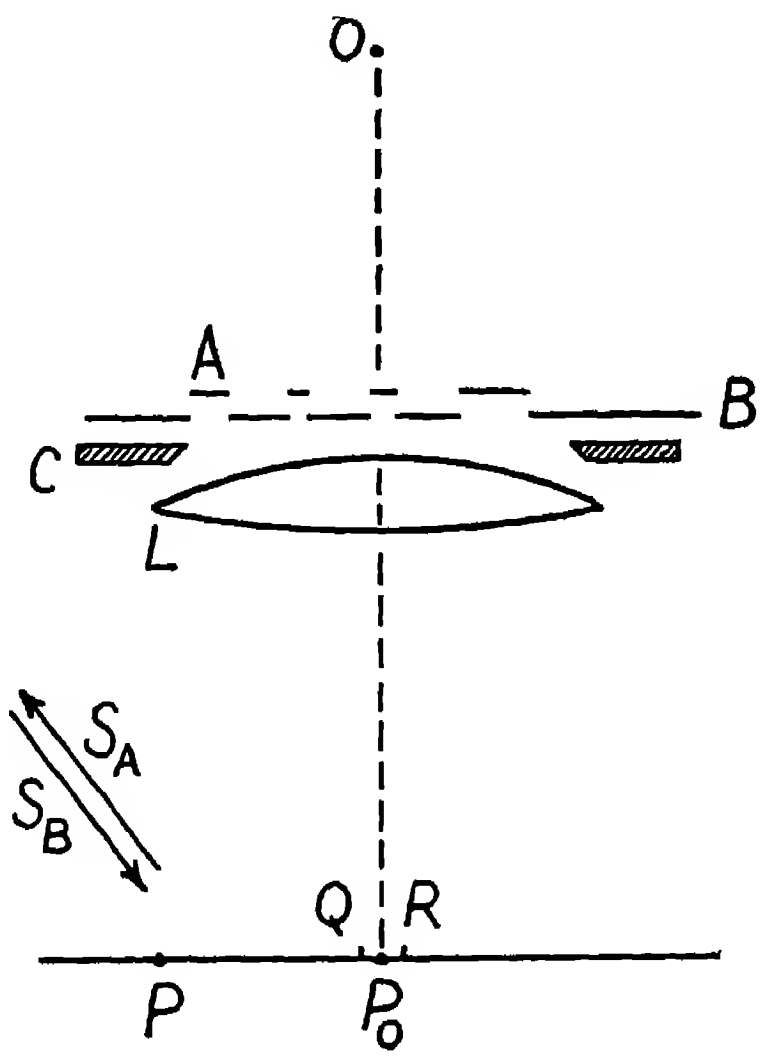


FIG. 181

Let O be a point source of light, L a lens, and P_0P the plane which is focally conjugate to the plane in which the source lies. When one screen alone is in place we shall have a Fraunhofer pattern in this plane. Now in the Fraunhofer arrangement, even when there is no diffracting screen present (screens A and B both withdrawn), the beam is still limited in aperture by the remaining optical system. The limit may be set by the periphery of the lens, but we may without loss of generality suppose that the limit is set instead by a diaphragm C having a large aperture which remains always in place.

When screens A and B are both withdrawn there will be an image of the source formed at P_0, or, more accurately, there will be formed a diffraction pattern of very small dimensions, which, practically speaking, will all fall

into a very small central region; let this region be represented by QR. The remainder of the field, lying outside of QR, will be dark.

When screen A is in place, B being absent, there will be illumination in various parts of the field. Let P be an arbitrary field point lying outside of the region QR, and let the vector S_A represent the resultant of the disturbances from the apertures in screen A which reach P. Since, when screens A and B are both absent, there is darkness at P, it follows that when screen B is in place, A being absent, the resultant disturbance from the apertures in B which reach P must be represented by the vector S_B, which is equal and opposite to S_A, because S_A and S_B must add up to zero. Consequently, the disturbance will be the same in magnitude with screen B as with screen A, and since P is any field point outside of the minute central region QR, we arrive at the startling conclusion that in Fraunhofer diffraction, complementary screens yield diffraction patterns which are identical everywhere except in a minute central region. The patterns are *identical*, and *not* complementary. For example, a small circular disk placed before the lens L yields exactly the same pattern as the complementary circular aperture—throughout the whole field except in the small central region, QR, occupied by the image which is formed when the disk and the screen with the complementary aperture are both absent. In this central region the illumination will be much greater with the disk than with the small aperture.

We are now in a position to understand a certain statement made in the chapter on the diffraction grating (chap. 6, sec. 15). The statement in question is that a grating composed of transparent strips of width b and opaque strips of width c, yields spectra of the same intensity as the complementary grating composed of opaque strips of width b and transparent strips of width c. The truth of this statement follows at once from Babinet's principle. The illumination will be the same with each of the two gratings everywhere except at the central image—the grating having the wider transparent strips will yield the more intense central image.

Passing now to the case of the Fresnel patterns which are formed by two complementary screens, let us consider first the instance in which there is a third screen having a large aperture, always in place. Referring to Figure 182, let O be the source of light, C the screen which remains always in place, and PP' a given field plane. This plane will be illuminated over a certain region when both of the complementary screens are absent. Let the geometrical cone of rays be indicated by the broken lines drawn from O by way of the edges of C to the field plane. There will be appreciable illumination somewhat beyond the base of this cone—let us say as far as the points Q and R. Beyond these points the field will be dark. When one of the two complementary screens is in place, call this screen A, there will be a diffrac-

tion pattern formed which in general will extend both over and beyond the region QR. When screen B is in place, A being absent, there will be a pattern formed which is different from that formed by A in the region QR, but outside of QR, for reasons explained in discussing the previous case, the pattern will be the same with screen B as with screen A. The present Fresnel case differs from the Fraunhofer case only in the fact that the region in which the patterns differ, the illuminated region QR, is now not minute, but is of relatively great extent.

We can draw certain conclusions in regard to the patterns formed in the illuminated region also; however, the conclusions which can be drawn for the illuminated region are somewhat scant. But such as they are, they apply equally well to the present, Fresnel, case and to the previous, Fraunhofer, case only in the Fraunhofer case the illuminated region, QR, to which the conclusions in question apply, is much smaller. The conclusions will now be deduced:

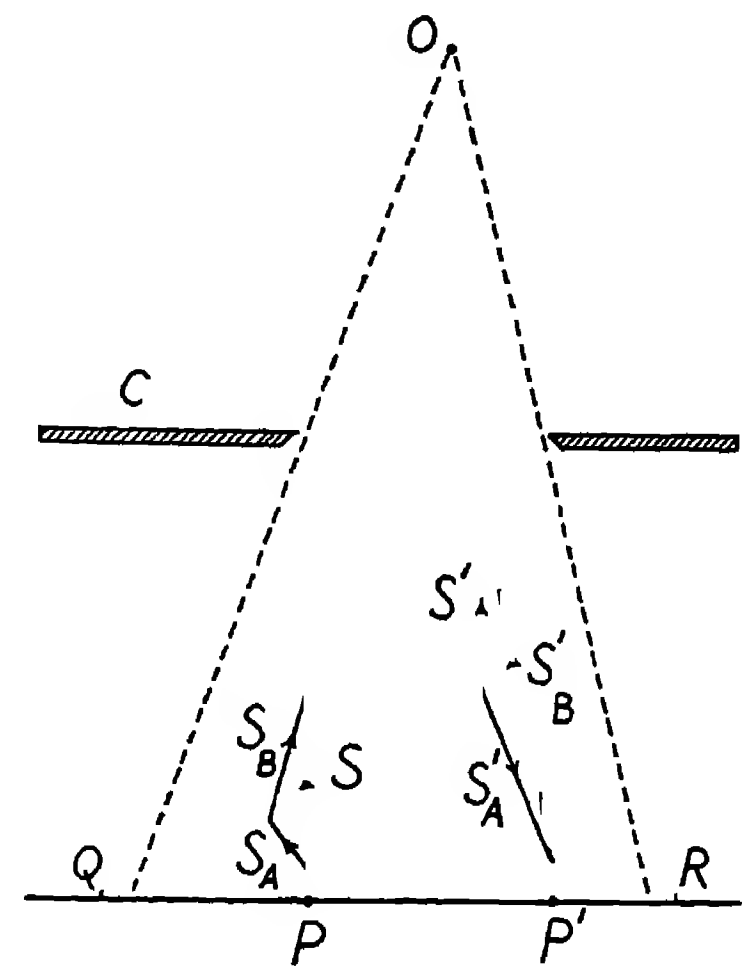


FIG. 182

The basic concept of Babinet's principle is that if the disturbance at any given field point P, when both of the complementary screens are absent, is represented by a certain vector S (Fig. 182), and the resultant disturbance from the apertures of screen A is represented by a certain vector S_A, then the disturbance from the apertures of screen B must be represented by a vector S_B which satisfies the vector addition:

$$S_A + S_B = S \tag{16}$$

If P happens to be a field point at which screen A yields no illumination, $S_A = 0$, then $S_B = S$. That is, the illumination at the field point in question, with screen B, must be equal to the illumination which obtains when both screens are absent. Conversely, at points where there is darkness with screen B, the illumination with screen A must be the original illumination. It does not follow, however, that at all points where the original illumination obtains with screen A that screen B will give no illumination. Let P' be a point at which screen A yields the original illumination. The corresponding disturbance may be represented by a vector

S'_A which is the same in length but differs in direction from the vector S', which represents the original disturbance at P'. The length of the vector S'_B, which represents the disturbance when screen B is in place, will now depend upon the angle which S'_A makes with S'. It is obvious that, depending upon the angle which these two vectors make with each other, the length of S'_B may take any value from zero up to the length $2S' = 2S'_A$—or, in other words, the intensity may take any value from zero up to four times the original intensity. In the same way, at points which have the original illumination with screen B we may expect to find with screen A an amount of illumination ranging anywhere from zero to four times the original illumination.

In the Fresnel case in which we deal with two complementary screens only, that is, when there is no third screen having a large aperture, which remains always in place, the entire field plane will be uniformly illuminated when both screens are absent. The patterns due to one of the two complementary screens will then in general differ from the pattern due to the other screen in all parts of the field. For example, the Fresnel pattern due to a circular disk is quite different from the pattern due to the complementary circular aperture. The only conclusion which Babinet's principle now allows us to draw is that at the points in each pattern at which the illumination is zero, if there be such points, there will be normal illumination in the other pattern, and that at the points in each pattern at which there is normal illumination, we shall find in the other pattern an amount of illumination which may have any value from zero up to two times normal illumination in amplitude, or four times normal illumination in intensity.

Babinet's principle also applies when we focus on the diffracting edges of two complementary screens introduced in succession. Let us suppose that we are using a telescope, in the manner described in section 10. When neither diffraction screen is in place, the extra-focal image of the source is the only region in the field of vision in which there is illumination. If now we introduce one of the two complementary screens, A, the diffracting edges will be luminous along certain stretches. There is now illumination along certain loci in the field of vision which were originally dark. Consequently, according to Babinet's principle, when screen A is withdrawn and the complementary screen B is introduced, the same loci must again be luminous and with the same intensity as with screen A. Moreover, at each point of the field of vision the phase of the disturbance which prevails with screen B must be the opposite of the phase which prevails with screen A. Therefore, the light deflected along a given portion of the diffracting edges, where light is deflected, with screen B, must be retarded in phase by one-half period with reference to the light which is inflected, along the same

portion of the diffracting edges, with screen A, and vice versa. The same portions of the diffracting edges appear luminous with either screen. The effect in the field of vision differs only in the region occupied by the extra-focal image. The extra-focal image will be obscured, wholly or in part, by one screen, and will be passed, wholly or in the same part, by the complementary screen.

Babinet propounded the principle which now bears his name in the year 1837. He discussed it only very briefly in a memoir on meteorological optics.* We shall learn later why the principle is important in this branch of science.

13. The Diffraction of Light by Small Particles.—Considering first the complementary case of diffraction by small apertures: Let us suppose that we have a point source of monochromatic light and that we focus the eye upon this and then place immediately before the eye an opaque screen with a single minute circular aperture in it. Under these conditions the Fraunhofer pattern due to the circular aperture is formed upon the retina of the eye, and accordingly one sees, in space, a circular central maximum surrounded by one or more rings or halos. The point source of light will not be seen as such, in focus. The smaller the aperture is, the larger will be the pattern, in the same proportion. A simple calculation shows that when the aperture has, for example, a diameter of .001 cm (10μ), the pattern will occupy a field of vision having an angular semidiameter of some 10°.†

Let us now consider an opaque screen having numerous small apertures all of the same diameter. Let the apertures be close together and at random positions. Such a screen when held before the eye will yield the same pattern upon the retina as that formed by a single aperture, except that the pattern formed by the numerous apertures will have an intensity proportional to the number of apertures which are effective, and moreover the point source of light will be seen in focus at the central point of the resultant pattern. That the pattern will take the form described follows from the fact that when the apertures are spaced at random there can be no observable interference effects between the disturbances from the various apertures, each considered as a unit, except at the central point of the pattern. Here the disturbances from all of the apertures arrive in the same phase with

* A. Babinet, *Compt. rend.*, **4**, 638, 1837; see in particular bottom of p. 643 ff.

† Assuming that the central maximum and the first two halos and these only will have sufficient intensity to be observable, it will be pertinent to calculate the semidiameter of the third dark ring. Referring to chap 5, sec. 3, we have, according to Table II, for the third dark ring $\delta = 3.24\lambda$. The corresponding angular semidiameter is hence $\theta = 3.24\lambda/d$, where d is the linear diameter of the aperture. Hence with $\lambda = 5 \times 10^{-5}$ cm, and $d = .001$ cm, $\theta = 3.24 \times 5 \times 10^{-2}$ radians $= 9.3°$.

the result that there is constructive interference at this point which leads to the formation there of a well-focused image of the source.*

A diffraction screen which is the complement of one having numerous small apertures all of the same diameter may be formed by depositing minute disk-shaped or spherical particles of uniform size upon a glass plate. Babinet's principle now tells us that such a screen held before the eye must yield the same pattern as that formed by circular apertures having a diameter equal to that of the particles. The pattern must be the same throughout the whole field except at the central point. Here a well-focused image of the source will be formed with either a screen of apertures or with one of particles, but the intensity of the image of the source will be different with one screen from what it is with the other. The intensity of the image formed at the center will depend upon the total area which is transparent, with the screen in question.†

Our discussion thus far applies to apertures or particles which, while small, nevertheless have a diameter which is at least several times as great as the wave-length, and we shall for the present discuss only the effects produced by particles having a diameter of this order of magnitude. The diffraction effects produced by particles which have a diameter about equal to the wave-length, and the effects produced by particles of which the diameter is much smaller than the wave-length, each require separate discussion. This will follow in due course.

14. Young's Eriometer.—This is a simple instrument used by Thomas Young‡ for determining the diameter of small particles of uniform size having a diameter of at least several wave-lengths. The diameter in question is determined by measuring the angular diameter of one or the

* In the same way numerous narrow slits of equal width, spaced at random, yield a pattern which is the same as that due to any one of the slits, only more intense, and, moreover, the source is seen in focus. In contradistinction, numerous *equally spaced* slits form a grating, and this yields a pattern in which there is complete constructive interference at certain points and practically complete destructive interference at all other points. But when sufficiently numerous slits are spaced at *random*, neither constructive nor destructive interference will be observable, except at the center of the pattern where the interference is constructive.

† The amplitude will be proportional to the transparent area, and hence the intensity will be proportional to the square of this. The sum of the intensities of the two images formed, one with each screen, will *not* give the intensity of the image of the source which is formed when both screens are absent; the two screens each diffract light into outlying parts of the field, and when both screens are absent this light also goes into the image. Consequently the intensity of the image formed when both screens are absent is greater than the sum of the intensities of the two images formed when the two complementary screens are used in succession.

‡ *Miscellaneous Works*, **1**, 305 and 346.

other of the halos which the particles produce. The eriometer may also be used for measuring the diameter of thin fibers. In fact, Young devised the instrument for the purpose of grading specimens of wool in regard to fineness or coarseness of fiber. The prefix "erio" is derived from the Greek ἔριον, meaning "wool." The fibers do not need to be straight or parallel. It is in fact preferable to leave them oriented at random, and when so oriented they, like small particles, produce circular diffraction maxima.

The eriometer consists primarily of a metal plate having a hole of about .5 mm in diameter at its center, and, surrounding this central hole, at a radius of 10 or 15 mm, a ring of much smaller holes. The entire system of holes is backed by some primary source of light of sufficient size. The central hole serves as the main source of light. The smaller holes in the ring give only enough light to permit this ring to be used as a reference mark against an otherwise dark background. The specimen to be examined is mounted in a support, and the observer brings his eye up close behind the specimen and looks through it at the central hole in the plate, whereupon he sees this hole surrounded by one or more halos. He now varies the distance of the specimen from the plate, keeping his eye always close behind the specimen. The position of the specimen may be read off on a linear scale provided for this purpose. The halos each retain a fixed angular diameter throughout, but when the distance of the specimen from the metal plate is made, let us say smaller, the apparent *linear* diameter of each halo projected against the metal plate becomes smaller in the same proportion. Now the angular diameter of for example the first halo is large, in the same proportion in which the particles which are under examination are small. Hence, the smaller the particles, the closer must the specimen be brought to the plate in order to reduce the apparent linear diameter of the halo to that of the ring of small holes. And this is what is done in the process of measurement. The distance of the specimen from the plate is adjusted until for example the first halo has the same diameter as the ring of small holes. Then the distance of the specimen from the metal plate is read off on the above-mentioned linear scale. The diameter of the particles or fibers is in direct proportion to the distance of the specimen from the plate. This diameter may be ascertained either by calculation, or, perhaps more accurately, by calibrating the instrument through the use of particles or fibers of which the diameter has been previously determined microscopically.

The eriometer is simple in construction, easily handled, and presents a ready means of obtaining the average diameter of small particles or fibers which all have nearly, but not accurately, the same diameter. By means of it Young investigated the diameters of a long list of particles and fibers having diameters ranging from about 3 μ to about 50 μ. The corpuscles of

human blood were found to have a diameter of about 7μ. Lycopodium powder gives the most distinct halos indicating that the particles are very nearly uniform in size. The diameter of the particles of this vegetable powder is about 27μ, or about 54 wave-lengths.

15. Meteorological Phenomena.—The science of meteorological optics deals with the numerous optical phenomena which are produced by the earth's atmosphere and by the solid or liquid particles which may upon occasion be suspended in it. Several of these phenomena involve diffraction, and we shall speak briefly of these and of others closely related to them. Further information will be found in the sources listed in the footnote.*

Halos, coronas, etc.—The well-defined ring of light, of large diameter, which is sometimes seen around the moon when the atmosphere is not entirely clear, is formed by refraction through alternate faces of small hexagonal crystals of ice which often float in the upper layers of the atmosphere. The ring in question invariably occurs at an angle of about 22° from the moon. Now, alternate faces of a hexagonal crystal form a truncated prism of 60°, and when light passes at minimum deviation through a prism of ice of angle 60° it is deviated by 22°. Accordingly, it is safe to suppose that the 22° ring is due to such refraction. It is not necessary that all of the ice crystals be oriented in any special manner. Light is refracted to the eye of the observer, at minimum deviation, by those crystals which happen to be in about the proper orientation. In the vicinity of minimum deviation, as is familiar, a comparatively wide latitude is permitted in orienting a prism. Moreover, since the angle of deviation involved is the minimum possible one, the inside of the ring is rather sharply defined. Crystals located in the sky at angles greater than 22° may also refract light to the eye, but since, for these, refraction must take place at an angle of deviation greater than the minimum one, the requirements of orientation will be more exacting and consequently there will be fewer crystals participating in refraction at a given larger angle at a given instant. Accordingly, the ring fades off gradually, and yet fairly rapidly, toward the outside. Moreover, since red light is refracted at the least angle of all of the colors, the ring will be somewhat red in hue on the inside—and possibly somewhat blue on the outside.

The 22° ring may also be seen around the *sun*. In observing this ring or others around the sun, one should, however, view the portion of the sky which one desires to examine, through a filter or by reflection in still water in order to avoid direct sunlight, which is injurious to the eye.

A faint ring of light which is formed at an angle of 46° from the sun or the

* Pernter and Exner, *Meteorologische Optik* (1910); or Geiger and Scheel, *op. cit.*, Vol. **20**, chap. 3; or Wood, *Physical Optics* (2d ed.), chap. 12; or, for the theory of the rainbow only, Preston, *Theory of Light* (5th ed.), chap. 20.

moon also originates from refraction by ice crystals. This ring involves refraction through an end face and a side face of the crystal. These faces make an angle of 90° with each other and consequently lead to a larger angle of minimum deviation.

Other phenomena such as the "parhelic circle" and "mock suns" are also attributable to ice crystals, but the formation of these requires that the air be wind-free. In this case the crystals orient themselves to a certain extent in settling through the atmosphere.

The various appearances due to ice crystals are, as we would expect, seen most frequently and in most beautiful perfection in the arctic regions.

The rings at 22° and 46° are specifically called "halos."

Other rings not due to ice crystals also occur, and these are specifically called "coronas." These arise from diffraction by the minute globules of water which constitute a mist or a cloud. There may be a pattern consisting of several rings. The pattern is of the same nature as that obtained with Young's eriometer, and the entire pattern is of considerably smaller angular diameter than either of the above-mentioned halos. The diameter of the pattern is greater in proportion as the size of the globules constituting the mist is smaller.

The distinction in meaning between the words "halo" and "corona" is not always strictly drawn; the words are often used more or less as synonyms.

The fact that globules of mist are transparent, instead of being opaque, need not be taken into account; it is permissible to regard the globules as opaque and we shall so regard them. A transparent globule or globe brings light which is incident upon it to a focus, and then spreads it into a wide cone, and the result of this refraction is simply to superpose diffuse light upon the pattern with which we are concerned. This refers to the light which is refracted by the globe or globule without internal reflection. When internal reflection occurs, we have the situation which occurs in the formation of rainbows. But rainbows do not concern us now because the various possible rainbows lie in portions of the sky quite far removed from the region comparatively near the source in which coronas are formed.

When small particles are deposited upon a plate of glass and the plate is held immediately before the eye, the pattern formed upon the retina is strictly a Fraunhofer pattern, whereas in viewing atmospheric coronas the situation is more nearly that previously described in discussing the luminosity of diffracting edges. The observer is focusing on the diffracting screen, which is in this case the multitude of minute globules. If instead of dealing with globules we were dealing with globes large enough to subtend, each one, an appreciable angle at the eye, we would observe each globe to be luminous along a short stretch of its near edge, where the light from the source is

*de*flected, and also along the far edge, where the light from the source is *in*flected. If the globe were large enough, the eye would see these two luminous stretches separately—it would resolve them. Luminosity would be observed along these two stretches irrespective of the particular angular distance from the primary source at which the globe is observed. But since we are actually dealing with minute globules, the two luminous stretches are not resolved. If they are to produce any effect at all upon the retina, they must produce a joint effect. And they will produce an effect when the disturbances interfere constructively, and they will fail to produce an effect when the disturbances interfere destructively. Let us think in terms, not of the globule itself, but in terms of the complementary aperture. The observer will obtain an effect from the aperture if he is situated in that direction leaving the aperture, in which, for example, the first-order lateral maximum of the Fraunhofer type is formed. (We may speak in terms of the Fraunhofer pattern because the distance to the field plane in which the observer is situated is vast in comparison to the diameter of the aperture.) In other words, the observer will obtain an effect from all apertures, or globules, located at certain proper angles with reference to his line of vision to the source, and he will fail to obtain an effect from apertures, or globules, which are located in certain other directions, namely, in the directions along which the minima of the Fraunhofer pattern occur. This explains why the corona is formed, and explains why, in accordance with observation, each ring subtends the same visual angle, with particles of a given size, which the corresponding ring subtends when particles of the same size are deposited upon a plate of glass and the plate is held immediately before the eye. The pattern obtained is in effect again the same Fraunhofer pattern as that obtained with Young's eriometer.

In order that the corona may be reasonably well defined, the globules constituting the mist must be all of nearly the same size. When the corona is well defined it shows color, being blue on the inside and red on the outside. This is what we would expect since long wave-lengths are diffracted to greater angles than short ones. By way of distinction, the halos formed by ice crystals are red on the inside and blue on the outside, as was stated above.

Coronas may also be formed by mists or dust clouds produced in the laboratory.

An observer standing on a mountain or wherever he may look at a sufficient angle downward, diametrically away from the sun, may, if there chance to be a cloud in that direction, see his shadow against the cloud, and may see a bright halo, a so-called "glory," surrounding the shadow of his head. This glory originates from light which is reflected and diffracted by the globules of mist which form the cloud. In observing the halo in ques-

tion, if there be several observers, each observer sees a halo encircling only his own head. To each observer the shadows of the heads of his companions appear to lack halos. This aspect of the phenomenon is particularly striking when the cloud is close by.

The rainbow.—We shall not give the theory, but content ourselves with making a few comments. The complete theory is both lengthy and difficult. It involves refraction of the incident light at the surface of the raindrop, internal reflection, another refraction upon emergence, and subsequent interference and diffraction effects. The theory may be found in numerous places, among others in the books referred to in the previous footnote. The gross characteristics of the principal rainbow and of the fainter rainbows which are sometimes seen are satisfactorily accounted for by considering only the refractions and reflections which take place at the surface of, and within, the raindrop. Since the raindrop is supposedly spherical, and all spheres are similar, this theory calls for the formation of for example the principal rainbow, always at the same angle, 138°, away from the sun, independently of the size of the raindrops giving rise to the bow. The semidiameter is supposedly always $180-138=42°$, with the center of the circle in the direction away from the sun. Actually, the semidiameter varies from occasion to occasion by several degrees, the bow is not always accurately round, the colors are not always equally pure, and there may be additional fainter bows close by the main one.

The theory which takes account of refraction and internal reflection only was first given by Archbishop Antonius de Domini in the year 1611 and more clearly by Descartes in 1637. This theory is now usually spoken of as Descartes's theory, and in this theory it is shown that for each color the light is most intense in the direction in which the minimum possible deviation for that color occurs—and this is the direction in which the color in question is seen.

Descartes's theory leads to the diameter of the bow which actually obtains when the raindrops are large. But when the raindrops are small the bow is slightly smaller. The lack of roundness of the bow which sometimes occurs is accounted for by the fact that the drops are of different size in different parts of the sky, and the occasional lack of purity of color is accounted for by the fact that the drops in one part of the sky are not all of approximately the same size. In the year 1838 G. B. Airy* developed a more complete theory which takes account of the additional effects which arise from interference and diffraction. These effects become more and more marked as the drops become smaller, and it is these effects which account for the various anomalies which we have mentioned.

* *Trans. Camb. Phil. Soc.*, **6**, 379, 1838.

The blue color of the sky.—The fact that we receive subjectively blue light from a clear sky is the result of so-called "Rayleigh scattering." This type of scattering will form the subject of the section after the next, and there, among other things, the reason for the blue color will be discussed.

16. Quetelet's Rings.—Suppose that a plane mirror be lightly dusted on its front, unsilvered surface, with, for example, lycopodium powder, and a small source of light held close beside the eye be then viewed in the mirror, the mirror being placed at a distance of several feet. Under these circumstances one sees a set of halos similar to those obtained with Young's eriometer or in atmospheric coronas, and superposed upon this system is a set of rings which have come to be known as "Quetelet's rings."

When the light incident upon the front surface of the mirror is diffracted by a given particle, it then passes to the rear, silvered surface of the mirror, is here reflected, and, in subsequently passing out through the front surface, is again diffracted. In any given direction of emergence the two disturbances which arise from the two successive diffractions by the same particle are in a condition to interfere. As a result a system of rings will be formed which have the normal to the mirror as a common center.

These rings were first observed, in the manner described, by Dr. Whewell, who communicated the information to Quetelet who made it public. Somewhat similar rings were investigated by Newton. He used a concave mirror, in a different manner. The rings then formed are known as "Newton's diffusion rings." For further information regarding the formation of rings of this type see the references given in the footnote.*

17. Rayleigh Scattering.—We have been considering diffraction by small particles, but the particles in question while small have yet supposedly had diameters several times as great as the wave-length. We shall presently consider diffraction by particles which are very much smaller still. We may throughout think of the particles as being spherical.

Speaking in terms of the circular aperture which is complementary to a particle of any given size: A circular aperture which has a diameter of at least several wave-lengths will yield a pattern which is characterized by the presence of maxima and minima. These maxima and minima may be thought of as being those of the Fraunhofer pattern, for we shall continue to concern ourselves only with cases in which the distance of the source and the distance of the field point are vast in comparison to the diameter of the aperture. The smaller the diameter of the aperture, the larger is the whole pattern and, in particular, the broader is the central maximum. If we consider the diameter of the aperture to decrease continually, we shall eventu-

* Preston, *op. cit.* (5th ed.), p. 233; Wood, *op. cit.* (2d ed.), p. 242; Winkelmann, *Handb.* (2d ed., 1906), **6**, 1083; Raman and Datta, *Phil. Mag.*, **42**, 826, 1921.

ally arrive at a situation in which the central maximum just extends over all possible angles of diffraction, from zero to 90°. This situation will occur when the path increment at the edge of the aperture which is required for the formation of the first minimum, namely, the path increment $\delta = 1.22\ \lambda$, can be obtained only by going to the extreme possible angle of diffraction, namely, 90°. The diameter of the aperture which then obtains is $d = \delta = 1.22\ \lambda$. For an aperture of this diameter, and for all smaller apertures as well, the only diffraction which we can have is a general diffusion of illumination, so-called "diffuse scattering." There are, however, certain considerations (which are explained in sec. 25) which make it impossible to rely fully upon diffraction theory in the form thus far presented when the aperture, or obstacle, has a diameter of only the order of magnitude of the wave-length. Accordingly, the aforementioned factor 1.22 has really no significance. We may say, however, that diffuse scattering will set in when the aperture is reduced to a diameter somewhere in the neighborhood of one wave-length and will continue for all smaller apertures.

Invoking Babinet's principle, we may now make a statement similar to the foregoing in regard to small particles: Particles having diameters of the order of the wave-length and all smaller particles as well will produce only diffuse scattering.

The particular case which we shall next consider is that of diffraction by particles of which the diameter is only a *small fraction* of the wave-length. In this case the path differences from various points of the particle (or from various points of the complementary aperture) to any given field point are all so small that the corresponding phase differences need not be taken into account. The question of phase differences does not enter the problem.

The principles which govern the scattering of light by particles which are small compared to the wave-length were first discussed by Lord Rayleigh in the year 1871. It is the scattering from such particles which has now come to be known as "Rayleigh scattering." Lord Rayleigh returned to the discussion of the question upon two occasions in later years.*

Referring to Figure 183, let the light be incident horizontally from the left as indicated by the arrow A. Let this light be plane polarized, the direction of the vibration being vertical. This refers to the direction of vibration of the ether particles when we are speaking in terms of the elastic solid theory, or to the direction of the electric vector when we are speaking in terms of the electromagnetic theory. Let D be a diffracting particle which is small

* *Op. cit.*, **1**, 87, 104, and 518 (in particular, pp. 528 ff.); **4**, 397. These papers all appeared in the *Phil. Mag.* for the years 1871, 1881, and 1899.

For a recent discussion of Rayleigh scattering see Jean Cabannes, *La diffusion moléculaire de la lumière* (Paris, 1929).

compared to the wave-length. The particle may be supposed to "load" the ether. However, if we could apply to the ether at this point a proper periodic force, in addition to the force which is applied by the incident light, and having the same frequency, the displacement in the ether could be everywhere and at every instant restored to what it would be if the particle were absent. In this case there would be no scattering. But since the additional periodic

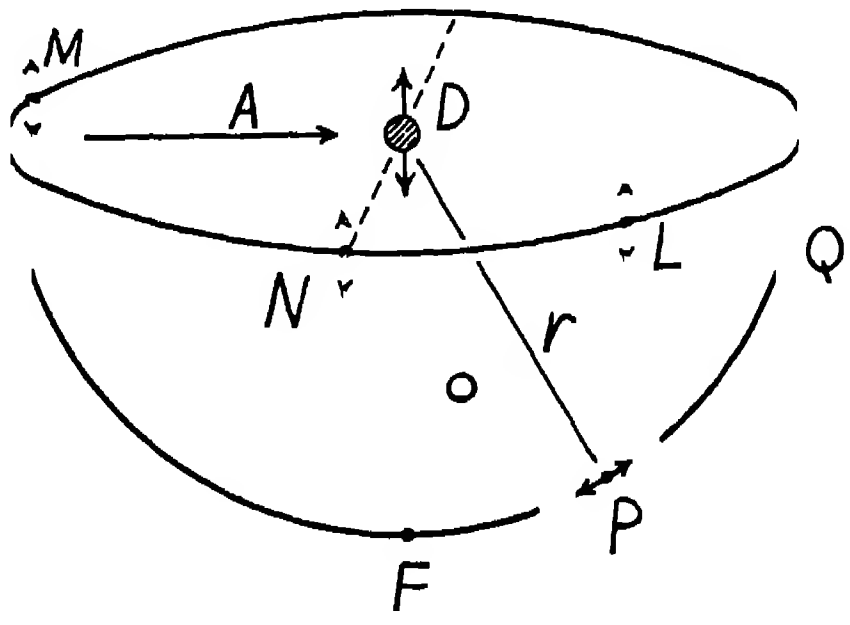


FIG. 183

force is not forthcoming, the particle will act as a center of scattering. Now it follows from considerations of symmetry that the additional force would have to be vertical, and consequently the scattering effect produced by the particle will be the same as though, in the absence of the primary wave, we applied the proper vertical periodic force at the position of the particle. This force would have to be at every instant the negative of the above-mentioned additional periodic force which would be required to neutralize the effect of the presence of the particle when the wave is passing.

The scattering particle is not actually set into bodily motion by the incident light, but the electrons within the atoms which constitute the particle are set into vibration. The particle constitutes a minute linear oscillator which is undergoing forced vibration.

Let us consider the disturbance which the vibration at D causes at various points on a reference sphere which has the point D as center. Let LMN be a horizontal circle described on this reference sphere and let FPQ represent a vertical circle described thereon. Denote the radius of the sphere by r. At all points on the horizontal circle—for example, at L, M, and N—the disturbance will be vertical and will have everywhere a common amplitude. Denote this amplitude by a_0. Continuing, let F be a point along the vertical line through D. The vibration at D has no component perpendicular to the direction of propagation from D to F; consequently the disturbance at F will be zero. Further, at the field point P the amplitude will be $a = a_0 \sin \phi$ where ϕ is the angle indicated. This result follows from the fact that that

component of the vibration which obtains at D which is perpendicular to DP is obtained by multiplying by $\sin \phi$. The *direction* of the vibration at P, or at any other field point, is determined according to the principle that the vibration there must be perpendicular to the line of propagation to the field point and must furthermore lie in that plane which contains the line FD along which the vibration at D takes place. When the field point is considered to rotate about FD as axis, the disturbance remains constant in amplitude but varies in its direction in space. The direction is throughout determined according to the principle which was just stated. It is to be noted that the direction which the vibration takes is such as to make the angle between this direction and the vibration at D a minimum, or, what is the same, to make the angle with the direction of vibration of the primary light a minimum.

We should of course think, not in terms of a single particle, but in terms of a multitude of particles suspended in a gas or a liquid contained let us say in a bulb, and we should then consider what we will observe when we take up various positions around this bulb.

Scattering experiments serve to answer what was long an unanswered question, namely: What is the direction of the vibration in a beam of light which has been plane polarized by refraction through a crystal or by reflection from, for example, glass or water? When the light incident upon scattering particles has been plane polarized by passing it through a nicol prism, the nicol must have the long diagonal of its end face horizontal in order that the scattered light may have constant intensity in the horizontal plane, and zero intensity vertically, as represented in our figure. Accordingly, the direction of vibration in the light transmitted by a nicol prism is vertical when the long diagonal of the end face is horizontal. Now a nicol prism consists of a crystal of calcite of certain proportions which has been cut into two pieces, and cemented with Canada balsam; the layer of Canada balsam reflects the ordinary ray to the side of the prism where it is absorbed, and accordingly the nicol transmits the extraordinary ray only. It is thus the extraordinary ray which furnishes the light when we allow light polarized by a nicol prism to be incident upon scattering particles. And since the long diagonal of the prism must be horizontal when the vibration in the scattered light and hence also that in the incident light is to be vertical, we arrive at the conclusion that the direction of vibration in the extraordinary ray is the direction perpendicular to the long diagonal of the end face of the crystal.

Unfortunately, long before there was any way of knowing which was the direction of the vibration, the plane which is *parallel* to the long diagonal of the crystal came to be defined as the "plane of polarization" of the extraordinary ray, and this definition still stands. In other words, what is by

definition the "plane of polarization" is perpendicular to the plane in which the vibration takes place. The latter plane is now called the "vibration plane."

It turns out, in similar fashion, that when light is plane polarized by reflecting it from glass or water at the proper angle, the direction of vibration is parallel to the surface; accordingly, the "plane of polarization" is perpendicular to the surface or in other words coincides with the plane of incidence and reflection.

Before there existed any direct experimental evidence indicative of the fact that the vibration plane is perpendicular to what had been defined as the plane of polarization, Fresnel, in arriving at certain equations for reflection and refraction at the surface of a transparent medium, which he deduced theoretically, had provisionally concluded that the vibration must be perpendicular to the arbitrarily defined plane of polarization. There now also exist other experiments, besides those of scattering, which all lead to this conclusion.*

The case in which the light incident upon the diffracting particles is unpolarized is also of importance. In order to comprehend the results which then obtain we must think of the incident light as still being plane polarized but with the plane of polarization changing frequently, or let us say rotating about the direction of incidence as axis, that is, rotating about the vector A of our figure. When the incident light is unpolarized and we then observe along the direction of incidence, or practically along this direction, as from the point M of our figure, we find the diffracted light unpolarized. When, however, we observe from N we find the state of polarization to be the same as when the incident light is polarized with the direction of vibration vertical, because, in observing from N, only the vertical component of the vibration at D is effective. In observing from F we shall now find the same intensity as in observing from N and we shall find the diffracted light plane polarized with the vibration perpendicular to the plane of the paper. In observing from a general field point—for example, from P—we find partial polarization.

With particles too large to yield pure Rayleigh scattering, anomalies are sometimes found in regard to the state of polarization of the scattered light. In some cases the direction of vibration is at right angles to the expected direction.

Supposing that the light which we receive from the sky is due to Rayleigh scattering caused by minute particles of one type or another, we would expect, according to the principles of scattering which we have been consider-

* Notably those of P. Selenyi. For a summarized discussion and references see Wien-Harms, *Handb. d. exp. Phys.*, **18**, 135 and 284.

ing, to find the following: In looking toward the sky anywhere at right angles to the direction of the sun and examining with a nicol prism, we would expect to find the light completely plane polarized. In looking in other directions, we would expect to find partial polarization, except that if, at sunrise or sunset, we should look along the direction of the incident light, we would expect to find the light completely unpolarized. Actually several directions may be found where there is no polarization, and, moreover, in looking at right angles to the sun, while there is a considerable amount of polarization in this direction, the polarization is by no means complete. This lack of completeness may possibly arise from the fact that we are not dealing with pure Rayleigh scattering, but other causes are more probable. The scattering particles are probably not spherical, and, moreover, the light which reaches any point of the atmosphere does not all come directly from the sun, but some of it is light which after having struck the earth is diffusely reflected. Hence, in the case of the light which reaches us from the sky, the exciting light is not all incident in one direction.

Much of the instigation which led Rayleigh to his consideration of the question of scattering was furnished by a series of experiments which had been performed by John Tyndall. In these experiments Tyndall produced "clouds" by the action of certain vapors upon each other. Tyndall made careful observations upon the changes in the scattered light which took place while the particles slowly increased in size, observing the color, analyzing the light with a nicol prism, etc. He discovered a number of interesting phenomena. One of these phenomena, which concerns us here, is that when the particles are first formed and are extremely small, the light scattered by them is always blue. Subsequently, when the particles have become larger, the light is white, because ordinary diffuse reflection then begins to set in. The blue color which appears when the particles are extremely small suggests that extremely small particles scatter short wave-lengths to a greater extent than long ones. This is as we should expect since a particle of given size constitutes an obstacle which is larger relative to short wave-lengths than to long ones. However, the rate at which the scattering increases with decrease of wave-length is cause for surprise. Lord Rayleigh deduced the law which relates the scattering to the wave-length in a highly mathematical treatment. He, however, also pointed out how the law might be arrived at from general considerations of dimensions. We shall follow him in the latter method of attack:

The amplitude of the disturbance caused by a minute spherical scattering particle of given material will be proportional to the volume of the particle. Denote this volume by v and denote the amplitude of the incident wave by A. Further, denote the distance to a given field point by r and denote by

ϕ the angle between the radius vector to the field point and the direction of the vibration in the incident wave. Introducing a constant of proportionality, K, the amplitude a which is produced at the field point, by the light scattered by the particle, may now be written:

$$a = \frac{KAv}{r} \sin \phi . \tag{17}$$

Now a/A is a ratio of amplitudes and is hence a pure number. Hence Kv/r must be a pure number. But v/r has the dimensions of length squared. Therefore K must have the dimensions of the reciprocal of length squared. But the only length involved in the problem which has not been taken into account is the wave-length, λ, of the light. Hence we may place $K = k/\lambda^2$, where k is pure number, and this number may depend only upon the constants of the material which constitutes the particle—for example, upon the index of refraction. Writing k/λ^2 in place of K, equation (17) becomes:

$$a = \frac{kAv}{r\,\lambda^2} \sin \phi . \tag{18}$$

Or we may write for the intensity i of the scattered light:

$$i = a^2 = \frac{k^2 A^2 v^2}{r^2\,\lambda^4} \sin^2 \phi . \tag{19}$$

The important result of this equation is that the intensity of the scattered light varies inversely as the fourth power of the wave-length. This means that blue light should be scattered about ten times as much as red light.

Before the day of Tyndall many physicists, from Newton downward, taking their clue from the interference colors shown by thin films, had attributed the blue of the sky to reflection from some kind of films in the atmosphere, the films being supposedly of the proper thickness to reflect blue light. After Tyndall's experiments there was, however, no longer any need to resort to such artificial hypotheses. It then appeared almost certain that the color of the sky arises from the greater scattering of the shorter wave-lengths by particles of some kind, but there was as yet no clue as to whether the particles responsible for the scattering were minute foreign particles, such as dust particles, or whether the molecules of the air might themselves be the centers of scattering. However, while this question had to remain open, Rayleigh showed experimentally that the intensity distribution of light from the sky follows his above-given inverse fourth-power law as closely as would be expected. Rayleigh showed this by passing light from the sky

through a spectroscope and comparing the intensity which was obtained at various points of the spectrum with the intensity which was obtained at the same points of the spectrum when sunlight diffused by a sheet of white paper was passed through the spectroscope.

In the year 1899 Rayleigh succeeded in showing that the blue color of the sky could be satisfactorily accounted for on the supposition that the molecules of air themselves constitute the scattering centers, and it has since become an accepted view that most of the scattering is caused by air molecules. It is of course not claimed that minute suspended particles do not also contribute, but the fact that the color of the sky is the same over the sea as it is over land indicates that the air molecules must be responsible for most of the effect.

According to a theory of Smoluchowski and of Einstein, the blue color of the sky may be due, not to air molecules acting each individually, but to transient, very local density variations which according to the kinetic theory of gases are constantly occurring at every point within a gas. These local variations of density occur because the molecules of a gas are in rapid motion.

The theory of scattering as developed by Rayleigh presupposes particles which are transparent. In Rayleigh's development, the scattering is attributed to the relatively small difference in optical density which supposedly exists between the particle and whatever the surrounding medium may be.

J. J. Thomson subsequently developed corresponding theory for an opaque, perfectly reflecting particle.

When the diffracting particle has a diameter which is not small compared to the wave-length, but is of the same order of magnitude, new aspects of theory arise. In particular, resonance effects play an important rôle.

For further information and references regarding the various theoretical developments just mentioned see the sources cited in the footnote.*

In some types of scattering such as, for example, in that involved in the Raman effect, a change in the wave-length of the light occurs. The theory of scattering when change of wave-length occurs falls outside of the province of diffraction theory.

18. Diffraction by a Slightly Transparent Screen.—Let us suppose that a film of silver be deposited upon one-half of a plane-parallel plate of glass of good quality, the other half of the plate remaining clear, and the film terminating along a straight edge. The plate, or rather the film upon it, will serve as a diffracting screen, and supposing momentarily that the film is completely opaque, this screen will form the usual Fresnel pattern due to a

* Geiger and Scheel, *op. cit.*, **20**, 307 ff.; *Encykl. d. Math. Wiss.*, **5**, Part III, 513 ff.

straight edge, having external fringes and no internal fringes. But if, on the contrary, the film transmits a small fraction of the incident light, the external fringes will be practically unaltered, and in addition there will be internal fringes formed. These internal fringes may be regarded as arising from interference of the small amount of direct light transmitted by the film with the likewise small amount of light which is inflected in the usual manner.

19. Diffraction by a Thin Transparent Lamina.—Consider a thin transparent lamina bounded by a straight edge to replace the usual opaque screen. Let the incident light be monochromatic of wave-length λ. Further, let the lamina, as viewed from a field point beyond the screen extend from the edge indefinitely to the left. The left half of the incident wave will pass through the lamina and the right half will pass by the lamina. Denote the thickness of the lamina by e and the index of refraction of the material, for the wave-length λ, by μ. Then the increment of optical path which a ray passing through the lamina suffers because the lamina is present is $\delta = (\mu - 1)e$. To begin with, let the thickness of the lamina be such that this path increment equals $\lambda/2$.

The diffraction pattern to which the lamina gives rise may be deduced from the Cornu spiral and accordingly the reader should think in terms of this spiral and the various closing vectors which enter into consideration, in reading the following paragraphs.

At what would ordinarily be the edge of the geometrical shadow, the amplitude, instead of having a value of one-half of that of normal illumination, will now be zero. This follows because at this point the contribution from the left half of the wave front will be equal in magnitude to that from the right half, and, moreover, the contribution from the left half will be retarded in phase by one-half period due to the presence of the lamina—hence destructive interference results.

If the lamina were opaque, the only fringes which would be formed would be the usual external ones, and these fringes could be interpreted as arising from the fact that the light which is stopped by the opaque screen is subtracted from the uniform illumination which would prevail in the absence of any screen, the subtraction being made in accordance with the principle of interference, that is, vectorially. When the transparent lamina is in place, we may regard the same light as first being subtracted and then added in reversed phase, and this operation would lead to the same result as though only the subtraction had been made, but in double amount. Accordingly, the final result is that what would with an opaque screen be the external fringes, will remain, fringe for fringe, but the maxima will be higher and the minima will be lower. The fringes will be more intense. Moreover, there will be an exactly similar set of fringes within what would with an opaque

screen be the shadow. The two sets of fringes are symmetrically located with reference to the "edge of the geometrical shadow." It is evident that this follows when we bear in mind that retarding the left half of the incident wave by one-half wave-length with regard to the right half, produces the same effect as advancing the left half by the same amount with reference to the right half.

The same pattern as that which is formed by a lamina causing a retardation of $\lambda/2$ will be formed by laminae causing retardations of $3\,\lambda/2$, $5\,\lambda/2$, etc. Furthermore, laminae causing retardations of respectively λ, 2λ, etc., would evidently form no pattern whatever, if the edge of the lamina could be cut absolutely "clean."

20. Talbot's Bands.—If, when an observer is viewing a continuous spectrum through the eyepiece of a spectroscope, he introduces, for example, a microscope cover glass from the side of the blue end of the spectrum, between his eye and the eyepiece, until the cover glass covers about one-half of the pupil of his eye, he will see the spectrum traversed by a series of vertical bright and dark bands. If he introduces the cover glass in the same way but from the side of the red end of the spectrum, the bands will not appear. Speaking in terms of the prism spectroscope, the foregoing means that if the bands are to be seen, the cover glass must be introduced from the side of the base of the prism when it is introduced as described, between the eye and the eyepiece.

The cover glass, or other transparent lamina, such as, for example, a sheet of mica, may also be introduced between the prism and the telescope objective. When introducing the lamina here it is well to stop the aperture down with a rectangular diaphragm of suitable dimensions and to mount the lamina on the diaphragm in such a manner that it covers one-half of the rectangular aperture and the other half is left clear. The diaphragm with the lamina mounted upon it must now however be introduced so that the lamina extends over that half of the aperture which lies toward the vertex of the prism. If introduced so that the lamina lies toward the base of the prism, the bands will not appear.

The bands may also be readily found, perhaps most readily, when the telescope is dispensed with and the spectrum is viewed directly in the prism. The lamina must in this case be introduced from the blue end of the spectrum which now lies on the side toward the vertex of the prism.

The bands in question were discovered by H. F. Talbot in the year 1837 and he gave an imperfect explanation of them based upon interference between the beam passing clear of the lamina and the beam passing through the lamina. But Talbot's explanation did not account for the fact that the bands fail to appear when the lamina is introduced on the wrong side. The

complete explanation requires that interference and diffraction effects be taken account of jointly. Since, however, the complete explanation is rather involved and the phenomenon is not one of great general interest, let it suffice to give references to where complete treatments may be found.*

21. The Colors of Mixed Plates.—The colors in question are readily produced by the following process: A fairly large drop of egg albumen or of saliva is placed on a plate of glass and this is converted into a froth containing numerous small air bubbles by pressing another plate of glass against the drop and lifting it again, and repeating this operation as often as may be necessary. The two pieces of glass with the froth between them are now to be worked over each other in a translational rotational motion, being at the same time pressed together. During this stage of the process the plates should be held up, at a distance of a foot or thereabouts from the eye, against the dividing line between a light background, such as is furnished by a window, and a nearly dark background, such as the wall of the room, provided the wall be sufficiently dark. The colors will soon appear, and when they have once appeared, colors will also be found when the specimen is held entirely against the light background and also when it is held entirely against the dark background, and whether the specimen be held at a distance or close before the eye. Finally, it should be noted that the specimen does *not* show color when viewed by reflection.

The colors in question were discovered and described by Thomas Young.† He found, upon examining the mixed film under sufficient magnification, that the minute air spaces are by no means all of the same size, and hence he concluded that the colors are not the result of diffraction, like the colors of halos, but arise from interference between the light which passes through the liquid and that which passes through the air spaces. Upon this view the color should be determined entirely by the optical path difference $\delta=(\mu-1)e$, where e is the thickness of the film and μ is the index of refraction of the liquid, those wave-lengths being removed from the incident white light for which the path difference δ as above given is an odd number of half wave-lengths. While it is true that the color observed is mainly a function of the thickness of the film considered in conjunction with the index of refraction of the liquid, Young's view does not completely account for the observed phenomena. The theory of diffraction by transparent laminae also enters. The colors shown by the so-called mixed plates, which are really quite thin mixed films, have been studied and treated by numerous investigators since

* Bouasse et Carrière, *Diffraction* (Paris: Delagrave, 1923), p. 128; Wood, *op. cit.*, (2d ed.), p. 254; Geiger and Scheel, *op. cit.*, **20**, 52; Winkelmann, *op. cit.* (2d ed., 1906), **6**, 1084.

† *Lect. Nat. Phil.*, **1**, 470; **2**, 635; *Miscel. Works*, **1**, 173.

Young's time. The complete theory is quite involved, and hence we shall content ourselves with giving references to where further information may be found.*

The mixed plate may also be composed of two different liquids. By way of example one of the liquids may be water, and minute globules of oil in the water may constitute the other liquid. Since the difference in index of refraction between two liquids is in general much less than the difference between air and a liquid, the thickness of the film which corresponds to a given optical path difference will in general be much greater when the film is composed of two liquids than when it is composed of air and a liquid.

22. The Action of a Grating on Single Pulses.—The question of what action a grating would have on isolated pulses arriving at random intervals is of interest in connection with the question of the nature of white light. What the connection is will appear as we proceed.

Referring to Figure 184, let AB be a single pulse having a plane front, which is traveling to the right, and is about to be incident upon the grating

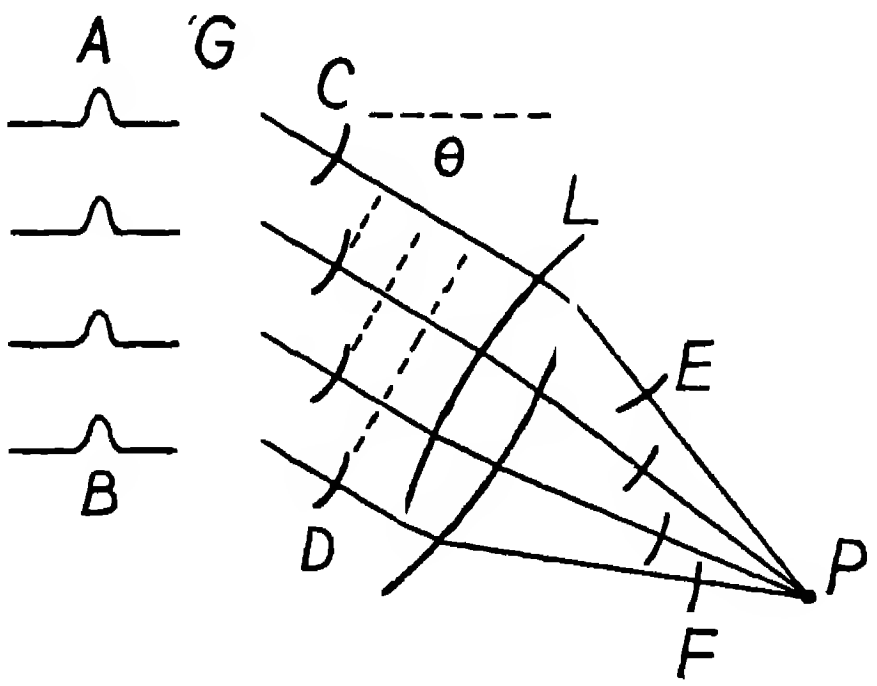


FIG. 184

G. For the sake of simplicity let us suppose that the direction of incidence is perpendicular to the grating, as represented in the figure. When the pulse reaches the grating each transparent strip of the grating becomes the point of origin of a Huygens wavelet and after a certain time has elapsed the wavelets will occupy the positions C–D; only a portion of each wavelet is represented. At any angle of diffraction θ, the wavelets follow each other at an interval which, measured along the given direction of diffraction, is equal to $\lambda = a \sin \theta$, where a is the grating space. The wavelets will pass through the focusing lens, L, and at a later time will occupy the positions E–F, and will eventually pass in succession through any given point in the focal plane of

* Raman and Banerji and Rao, *Phil. Mag.*, **41**, 338, 860, and **42**, 679, 1921; Wood, *Physical Optics* (2d ed.), p. 252; and Gehrcke, *Handb. d. Phys. Opt.*, **1**, 374.

the lens. Let P be the point in the focal plane which corresponds to the direction of diffraction θ. The wavelets will pass through P at the same space interval $\lambda = a \sin \theta$ at which they were following each other before passing through the lens. But P is the point to which a continuous train of waves of wave-length $\lambda = a \sin \theta$ would be diffracted. Hence we see, first, that a grating would manufacture out of a single pulse a disturbance which is *periodic* and, second, that the period of the manufactured disturbance, or its "wave-length," would be, at each point of the focal plane, the same as the wave-length of the continuous train of waves which would be diffracted to the same point. Each point of the focal plane would receive a manufactured periodic disturbance of the proper wave-length. Consequently, a succession of pulses arriving at random would form the same continuous spectrum which would be formed by trains of waves of all wave-lengths incident simultaneously upon the grating. The spectrum would in either case be identical with that which is formed by white light, and this fact forces us to entertain the alternate hypothesis that white light may consist of a random succession of pulses, instead of having trains of waves of all wave-lengths originally in it as we have hitherto supposed.

For the sake of simplicity the case of perpendicular incidence upon the grating was considered above, but obviously the same results apply at any angle of incidence.

A prism would act upon a random succession of pulses in a manner somewhat similar to that in which a grating would act upon them, but in the case of the prism it is possible to explain the action only after having considered the question of group velocity. We shall consider this question forthwith.

23. Group Velocity.—Suppose a stone is dropped into a quiet pond. Several successive waves will spread out circularly, and after these waves have proceeded some distance, a vertical cross-section of them will be somewhat as represented in the topmost curve of Figure 185. The waves constitute what we may call a "wave group." Let us suppose that the group is traveling toward the right. Observation of one of the crests near the front of the group—for example, of the one marked S—will reveal that this crest travels to the front of the group and disappears, and further obervation will reveal that each of the individual crests travels faster than the group considered as a whole, each crest traveling through the group from the rear to the front. If at one instant there is a crest Q at the center of the group, which has the greatest height, at a later instant the minimum between P and Q will be at the center of the group and the crests P and Q will have the same height. The group *changes* as time proceeds, and when the minimum between P and Q is at the center, the group may be considered to have become inverted. After a certain further time has elapsed the crest P will occu-

py the central position in the group and the group will then have returned to its original configuration.

Any wave group may be analyzed by Fourier series into a number of component wave trains extending to infinity in both directions. And this applies also when the group under investigation takes the form of a single pulse. We fortunately do not need to concern ourselves with the details of

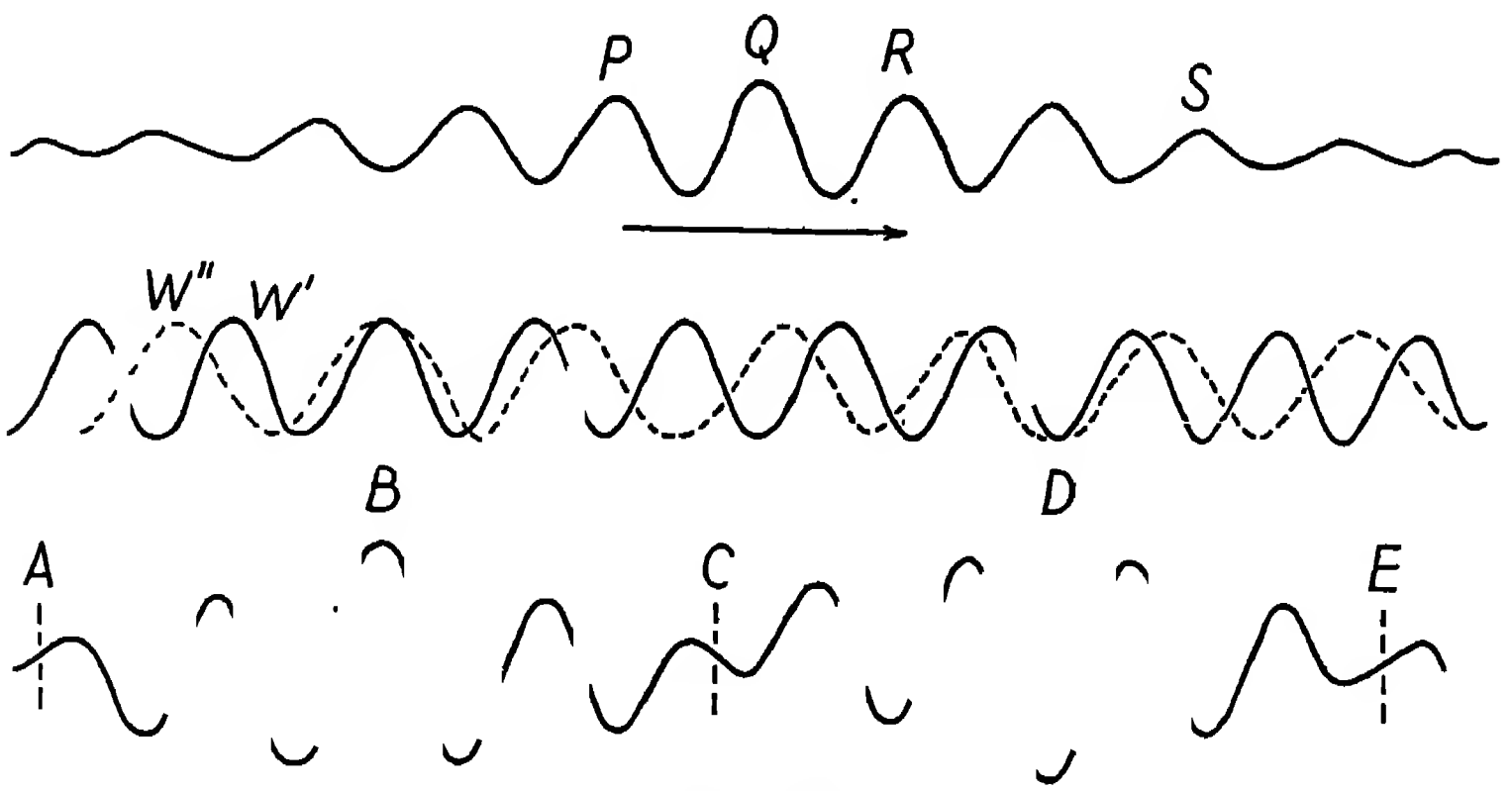


FIG. 185

the Fourier analysis but may learn all that we need to know about the propagation of wave groups by considering the summational effects which arise from only two continuous trains of waves of different wave-length traveling with different velocity. However, when there are only two component trains, the group as constituted at any one instant of time may repeat itself in space. This will become evident presently.

Referring to the middle and lower portions of Figure 185, let the solid curve W' and the broken curve W'' represent two continuous trains of waves of wave-lengths λ' and λ'', such that $\lambda'' > \lambda'$, and let the curve in the lower part of the figure represent the sum of the two curves above. In the illustration, two crests of the component trains coincide exactly at B and two minima coincide exactly at D. In this circumstance the group extending from A to C is repeated in inverted form between C and E. Beyond E the erect group will again be repeated and beyond this the inverted group, etc. Consider the wave trains W' and W'' both traveling to the right. In case the medium is one through which the two wave trains both travel with the same velocity, then each group of the summational train will obviously retain its form at all times; the summational wave will move to the right with the common velocity of the component trains, in its entire contour. The group velocity is then equal to the wave velocity—and there are no separate phenomena of group propagation which require consideration.

Taking up now the case in which the two trains travel with different velocities, let the velocity of the train W' be V' and let the velocity of the train W'' be V'', and, since $\lambda''>\lambda'$, let $V''>V'$, in accordance with the fact which obtains for an ordinary dispersive medium, such as glass, that the waves of greater wave-length travel with greater velocity. The velocities V' and V'' are known as "wave velocities" or "phase velocities" by way of distinction from the "group velocity" which will be defined shortly. When the particular crest marked W'' has caught up with the particular crest marked W', the original form of each group will be repeated identically, but each group will now be farther to the right by a distance which depends upon several factors. In the meantime the group will have passed through the inverted stage—but this fact does not now concern us directly. Let us fix our attention upon the original group AC. The crest W'' is overtaking the crest W' with a relative velocity $V''-V'$ and the distance between the two crests which must be closed up in order that the group may resume its original form is $\lambda''-\lambda'$. Denote by T the time which is required for the crest W'' to overtake the crest W', then:

$$(V''-V')\,T=\lambda''-\lambda'. \tag{20}$$

If the difference in wave-length between the two component trains is small we may write $\lambda''-\lambda'=d\lambda'$ and $V''-V'=dV'$, and equation (20) then yields

$$T=\frac{1}{\dfrac{dV'}{d\lambda'}}. \tag{21}$$

Since T is the time which elapses before the group next reappears in its original form, it may be called the "group period." Denote by X the distance to the right at which the group next reappears in the same form. Fixing our attention upon the train W' and in particular upon the crest B of this train, the distance X at which the next reappearance of the group will occur is the distance which this crest will have traveled in the time T less one wave-length λ'. Hence:

$$X=V'T-\lambda'=\left(V'-\lambda'\frac{dV'}{d\lambda'}\right)T. \tag{22}$$

We now define the "group velocity" as the distance which the group has traveled between identical reappearances divided by the corresponding time, or, denoting the group velocity by U, as:

$$U=\frac{X}{T}=V'-\lambda'\frac{dV'}{d\lambda'}. \tag{23}$$

The group velocity is *lower* than the wave velocity whenever $dV'/d\lambda'$ is positive—or, in other words, in any medium in which long waves travel faster than short ones. This is the condition which obtains in any ordinary dispersive medium such as glass or water.

Let us consider now the propagation of wave groups in a hypothetical medium in which the wave velocity is given by the equation:

$$V' = a + b\lambda', \tag{24}$$

where a and b are constants. We now have, when taking account of equations (21) and (23):

$$\frac{dV'}{d\lambda'} = b = \frac{1}{T}, \tag{25}$$

$$U = a, \tag{26}$$

$$X = UT = \frac{a}{b}. \tag{27}$$

For such a medium the space interval of repetition of a group, X, the group period, T, and group velocity, U, are all constant, that is, always the same no matter what the wave-lengths of the component trains may be, or how many component trains there may be, and hence always the same for all types of groups including single pulses. Let us now imagine that a medium of this type is formed into a prism, and imagine further that random pulses are incident perpendicularly upon one face of the prism.

24. The Action of a Prism on Single Pulses.—Referring to Figure 186, let random pulses having plane fronts be incident from the left upon the face AB of the prism ABC. Under these circumstances a given pulse will enter the face AB over its entire area at one instant of time. It will travel horizontally to the right, beginning to undergo change of form as soon as it enters the dispersive medium, the prism. If the pulse is an erect one to begin with, it will, at regular intervals, become inverted, will then become erect again, then again inverted, etc. Let the distance KD represent the distance previously denoted by X between successive recurrences of the same phase. The pulse will then crop out in its erect phase at the points D, E, and F of the face AC of the prism. If, as we shall suppose, the prism has a perfect edge at A, the point A should also be counted as a point where the pulse crops out in its erect phase. Halfway between the positions of appearance in the erect phase the pulse will crop out in its inverted phase. These intermediate positions are marked by small crosses. We shall, how-

ever, for the sake of simplicity limit the discussion to the appearances of the pulse in its erect phase at A, D, E, F, etc. At the instant at which the pulse enters the prism a Huygens wavelet may be considered as starting out from A. After a time T has elapsed the pulse will crop out in its erect phase at D and the Huygens wavelet from A may then be supposed to have arrived at

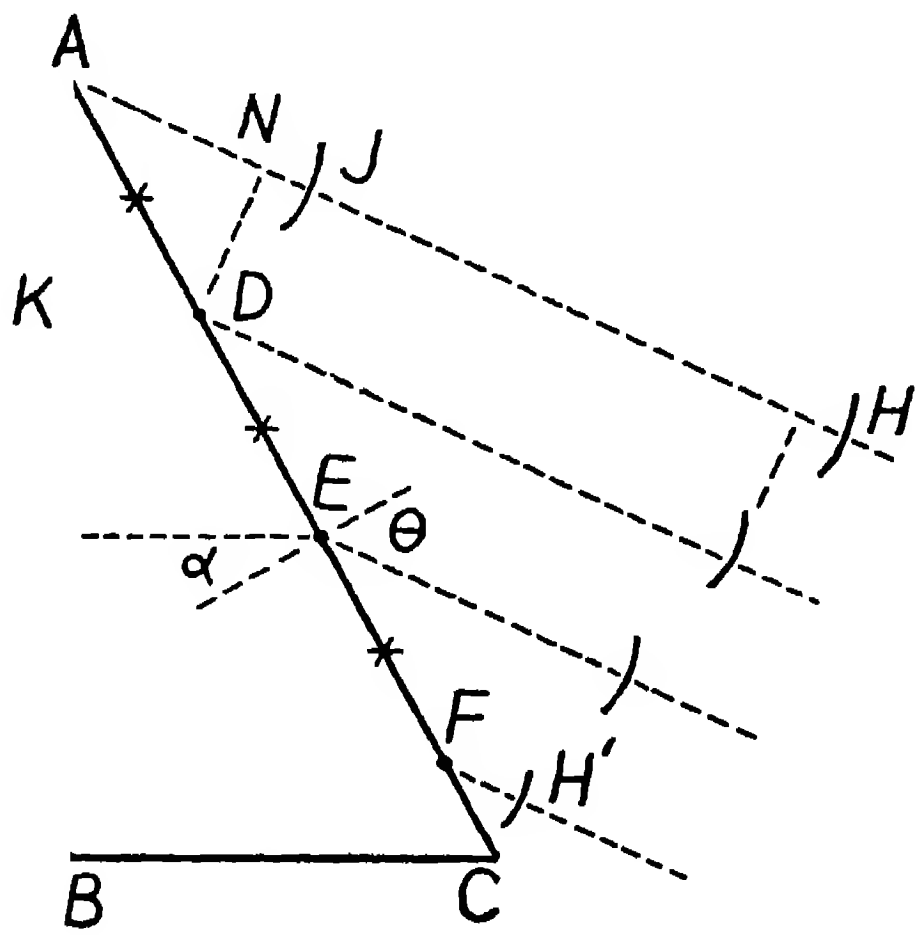


FIG. 186

a certain position J. A Huygens wavelet now starts from D, and at appropriate later instants similar Huygens wavelets start from E, F, etc. Let the wavelets in question be represented, at a later time, by H to H'. These wavelets must follow each other in the order indicated, and *not* H behind H' along any possible direction of propagation AH (or FH' parallel to AH); this follows because the group velocity is lower than the wave velocity and hence the wavelet H will be farther ahead than it would be if the group velocity were equal to the wave velocity, in which case the Huygens wavelets would all be abreast of each other. Furthermore, the wavelets obviously follow each other at equal intervals along any given direction of propagation and will consequently pass at equal intervals through any given point of the focal plane, when the prism forms part of a spectrometer, as we shall henceforth suppose that it does. Accordingly, a prism, as well as a grating, would manufacture a periodic disturbance out of a single pulse.

It remains to be proved that, at each point of the focal plane, the manufactured "wave-length" which the prism sends is the same as the wave-length of that continuous train of waves which will be refracted to the same point of the focal plane. Suppose that a given continuous train of waves has a wave-length λ in air or vacuum and a wave-length λ' within the prism.

Let this train be refracted at the second face of the prism at the angle indicated by θ in Figure 186. The value of θ will be determined by the equation:

$$\lambda = \lambda' \frac{\sin \theta}{\sin \alpha}, \tag{28}$$

where α is the angle indicated in the figure, that is, the angle of internal incidence upon the second face of the prism. This angle is, under the given condition of perpendicular incidence of the original beam upon the first face of the prism, equal to the prism angle. The Huygens wavelets which arise at the points A, D, E, F, etc., from a single pulse, follow each other, along the direction of propagation, θ, at a certain distance—denote this distance by λ_H. It is required to show that $\lambda_H = \lambda$.

We have:

$$\lambda_H = \overline{NJ} = \overline{AJ} - \overline{AN} \tag{29}$$

Denote the velocity of light in air or vacuum by c. Then $\overline{AJ} = cT$, where T is as before the group period or, in particular, the time which elapses between the instant at which the pulse enters the prism and the instant at which it appears at D. Moreover, $\overline{AN} = \overline{AD} \sin \theta$. Therefore:

$$\lambda_H = cT - \overline{AD} \sin \theta.$$

But $\overline{AD} = X/\sin \alpha$, where $X = KD$ is, as before, the interval at which the pulse reappears in its erect phase, and α is the prism angle which, as above pointed out, is under the given condition of incidence the same as the angle indicated by α in the figure. Hence:

$$\lambda_H = cT - \frac{X}{\sin \alpha} \cdot \sin \theta$$

$$= \left[cT \frac{\sin \alpha}{\sin \theta} - X \right] \frac{\sin \theta}{\sin \alpha}.$$

Substituting for $\sin \alpha / \sin \theta$, within the brackets, the equal quantity V'/c, we have:

$$\lambda_H = [V'T - X] \frac{\sin \theta}{\sin \alpha}.$$

According to equation (22), $V'T - X = \lambda'$, and hence:

$$\lambda_H = \lambda' \frac{\sin \theta}{\sin \alpha}. \tag{30}$$

This, when taking account of equation (28), brings us to the desired final result, namely:

$$\lambda_H = \lambda. \tag{31}$$

The same result holds when the direction of incidence of the pulse is not perpendicular to the first prism face, but the proof is then more difficult.

Continuing with the case of perpendicular incidence, we may consider a grating having a grating space which is equal to AD, DE, EF (Fig. 186), equivalent to our prism. With the beam incident upon the equivalent grating at the proper angle, the dispersion of the grating and of the prism would then be the same throughout if we could really find a prism having the hitherto supposed dispersive properties, that is, such that the wave velocity is given by the equation $V' = a + b\lambda'$. But this equation leads to a dispersion curve which is a straight line, and actual dispersion curves are not linear We may, however, consider any actual dispersion curve as straight over a sufficiently short wave-length range. Each of various short wave-length ranges will then lead to an equivalent grating having a different grating space, and in each range the dispersion of the prism will be the same as the dispersion of that grating which is equivalent for that range.

In an actual medium such as glass, the interval of group repetition X, the group period T, and the group velocity U, all vary as the wave-lengths of the component trains vary. Or rather, when there are numerous component trains, a given group, or, specifically, a single pulse, would never really repeat itself. But the action of an actual prism on single pulses would nevertheless be much in the manner which has been above described, with the difference that the action must be considered in terms of one short wave-length range at a time.

The formation of the continuous spectrum by either a prism or a grating may be as well explained by supposing that white light consists of random pulses as by supposing that it originally contains trains of waves of all wave-lengths. We are thus confronted with the question: In which of these alternative ways shall we consider white light to be constituted? A closely related question arose even before Newton's day, namely: Does refraction through a prism create the colors of the spectrum, or does it merely disperse colors originally present? Newton's experiments with prisms indicated that the colors were all originally present, and when the wave theory of light became established by the work of Young and Fresnel, from about the year 1825 onward, Newton's experiments came to be definitely interpreted as meaning that all wave-lengths must be originally present. But about a half-century later, when the realization dawned that a grating or prism

would manufacture a continuous range of wave-lengths out of random pulses, the question was asked in more modern form: May not white light consist merely of random pulses? For are not the conditions of chaotic temperature agitation which prevail in an incandescent solid much more favorable to the production of random pulses than to the production of trains of waves? Further, does not the fact that the spectrum due to white light reveals no "gaps" even when examined with a spectroscope of the highest resolving power favor the pulse hypothesis? The bearing of this last question upon the problem is as follows: With a given resolving power it is easy to calculate how near together the frequencies of different types of oscillators (electrons) executing sustained oscillations must fall if there are to be no gaps in the spectrum, and it is then easy to calculate how many different types of oscillators there must be. Now the number so calculated comes out vastly larger than can be reasonably accounted for. Hence it cannot be supposed that the oscillators execute sustained oscillations. So both of the foregoing questions were at one time answered in favor of the pulse hypothesis. But today it is further realized that which answer is to be given cannot be decided purely upon the basis of physical reality but involves questions of point of view as well. The disturbances which constitute white light must certainly lack obvious features of regularity, but we can resolve any given disturbance by Fourier analysis into component trains of waves of all wavelengths, which, taken collectively, are in every way equivalent to the original disturbance. Can we under these circumstances make a definite choice? However, if a choice be insisted upon, the pulse hypothesis probably corresponds the more nearly to physical reality.

While it is permissible to take the view that white light consists of pulses which arrive purely at random, we are obliged, in taking this view, to admit that there must exist controlling laws which determine the average time at which the pulses arrive, and the form which the pulses have, or at least their most probable form. The fact that a definite distribution of energy obtains in any given continuous spectrum requires that there be controlling factors. But the questions which arise in this connection pertain to the theory of radiation and they will not be entered into here.

While we are at liberty to choose our point of view with regard to the constitution of white light, monochromatic light must perforce be supposed to be due to a regular train of waves. Moreover, there must be more or less definite periodicity impressed even upon the light, covering a fairly broad spectral range, which is obtained when white light is filtered—for example, through colored glass.

The color which is observed in interference or diffraction fringes when such fringes are formed from white light should also be mentioned as being

among the phenomena which admit of explanation upon the pulse hypothesis.*

25. Critique of Fresnel's Theory of Diffraction.—The basic principles of Fresnel's theory were dwelt upon in chapter 1, section 5. According to Fresnel, the wave is to be regarded as being simply stopped, absorbed, over the area occupied by the diffracting system, and as passing, unaltered, over the remaining area; furthermore, Huygens' principle and the principle of interference are to be applied in conjunction with each other. It was stated in chapter 1, section 11, that the principles in question would carry us through the explanation of the main body of diffraction phenomena, and we have in the meantime seen that this is true. They have carried us through our treatment of diffraction phenomena of the Fresnel class, of the Fraunhofer class including those due to the grating, and, with some exceptions, through the treatment of the various phenomena discussed in the present chapter. It now behooves us to examine these principles critically:

The diffracting system regarded simply as a barrier.—Actually the diffracting system always reflects some of the light incident upon it, and for this reason, and for other reasons which will appear, there are limitations to the conditions under which Fresnel's theory may be applied. Referring to Figure 187, let light be incident from above in the figure upon a diffracting edge D, which we shall suppose to be a straight edge. Let P be a field point at a distance of perhaps a meter from the edge and a few millimeters within the geometrical shadow. Under these conditions Fresnel's theory applies, and we may determine the disturbance at P by taking the closing vector from the appropriate point in the upper branch of the Cornu spiral to the upper asymptotic point J. To the effective portion of the vibration curve corresponds the effective portion of the wave front, extending from D indefinitely to the right. Now it is a principle of Fresnel's theory that this portion of the wave front shall be considered as entirely unaltered. But it is only reasonable to suppose that actually there will be a restricted region near the edge where the wave front is affected by the presence of the diffracting screen. Let this region be DE and let the first half-period element of the wave front constructed from D to the right be DF. To the element DF corresponds the first effective semiconvolution of the vibration curve. It is ob-

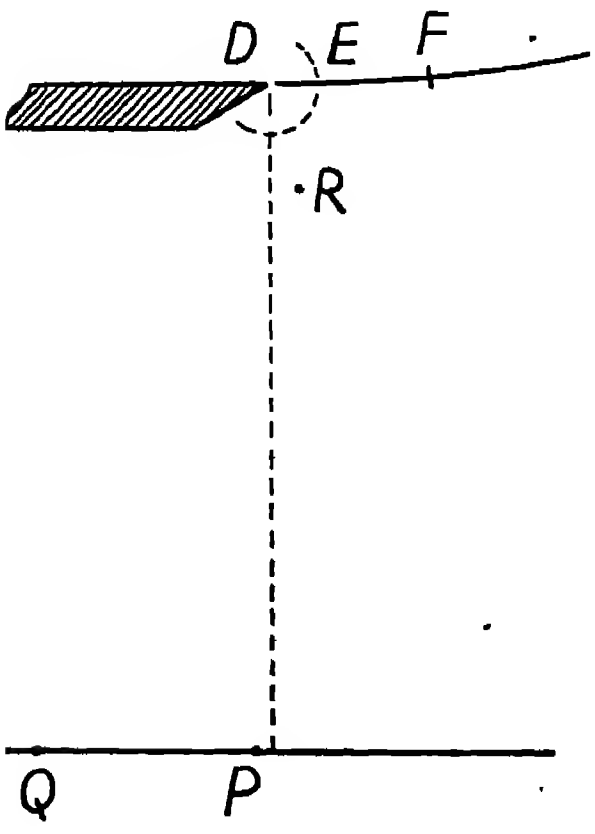


FIG. 187

* A. Schuster, *Phil. Mag.*, **37**, 509, 1894; or Wood, *Physical Optics* (2d ed.), p. 653.

vious that so long as DE is small compared to DF, Fresnel's theory should lead to a result which is correct to a high degree of approximation, but under any conditions under which DE is of the same order of magnitude as DF, Fresnel's theory may no longer be trusted to give even approximately the correct result. General considerations of wave theory lead to the conclusion that the distance DE, over which the wave is altered, has the order of magnitude of the wave-length, but the actual distance to which the alteration of the wave front extends in a given instance, and the degree of this alteration, depends markedly upon the material and form of the diffracting edge; Fresnel's theory neglects all effects of the material and form of the edge.

For a field point P lying, as we have supposed, at a distance of a meter from the diffracting edge and only a few millimeters from the edge of the geometrical shadow, the width DF of the first half-period element of the wave front will be large compared to one wave-length. Accordingly, under these circumstances Fresnel's theory should give the correct result to a high degree of approximation, and so it turns out to be. Let us now consider under what conditions Fresnel's theory might be expected to fail:

1. For a field point Q lying at a sufficiently large angle of diffraction, the width of the first half-period element of the wave front adjacent to the edge will be only of the order of magnitude of the wave-length. Hence Fresnel's theory should not be expected to hold at large angles of diffraction.

2. For a field point R close to the diffracting edge the first half-period element of the wave front will be only of the order of magnitude of the wave-length. Hence Fresnel's theory should not be expected to apply close to a diffracting edge. Moreover, when the field point lies close to the edge and at the same time outside of the geometrical shadow as illustrated, effects of ordinary reflection assume importance, especially when the edge is not sharp but is rounded (see chap. 1, sec. 1). On this account also the pattern would be different from that to be expected upon Fresnel's theory.*

3. When an aperture has a diameter of only a few wave-lengths, the altered region of the wave front extends over a relatively large portion of the aperture. Hence Fresnel's theory does not apply in this case. Similarly, his theory fails to apply to a screen which is complementary to such an aperture, namely, to an obstacle which has dimensions of only the order of

* Even a perfectly absorbing screen would cause an alteration of the wave front in a small region DE. When the screen is not perfectly absorbing, but reflects, we may for convenience speak of an additional effect of "ordinary reflection" in the illuminated region as a superposed effect. However, rigorously, there are not present two separate and distinct effects. But we shall not pursue this matter further, for in this connection arises the question of whether an "ideally black body" can exist, which is in itself a difficult and perplexing question to answer. For further discussion see, e.g., Geiger and Scheel, *op. cit.*, **20**, 286.

magnitude of the wave-length. Fresnel's theory is limited in its application to apertures or obstacles which are large compared to the wave-length.

We shall in later sections learn something of various comparatively recent investigations, both theoretical and experimental, which apply under circumstances in which Fresnel's theory is ruled out; we have, in fact, already made one excursion into this field in considering the phenomenon of Rayleigh scattering.

The application of Huygens' principle.—While Huygens' principle appeals to us as being eminently reasonable, and while it may, in a manner, be demonstrated experimentally, yet when we attempt to formulate this principle exactly, or attempt to apply it, we are beset with difficulties. Comprehensive consideration of the various difficulties which arise cannot well be entered into here. We must content ourselves with dwelling briefly upon a few of the more obvious and simple ones. If all that has been written upon Huygens' principle were collected, it would fill a large volume!

Let us first consider what difficulties arise when we attempt to apply Huygens' principle in a naïve manner to a single pulse. Referring to the upper part of Figure 188, let O be a source of wave disturbance, let AB be the locus of a single pulse having the form of a crest, and let P be a field point. According to Huygens' principle, we are to regard every point of AB as the source of a new wave crest. Let the resulting wavelets at a later time be represented by the small circles C–D. These wavelets should spread out in all directions, and hence there should be not only disturbances advancing toward P, but also disturbances propagated backward—there should be a so-called "back wave." Moreover, the wavelets should pass *in succession* through any given field point, for example, through P, and hence, instead of having a single well-defined crest pass through P, as actually occurs, there should be a feeble prolonged disturbance at P.

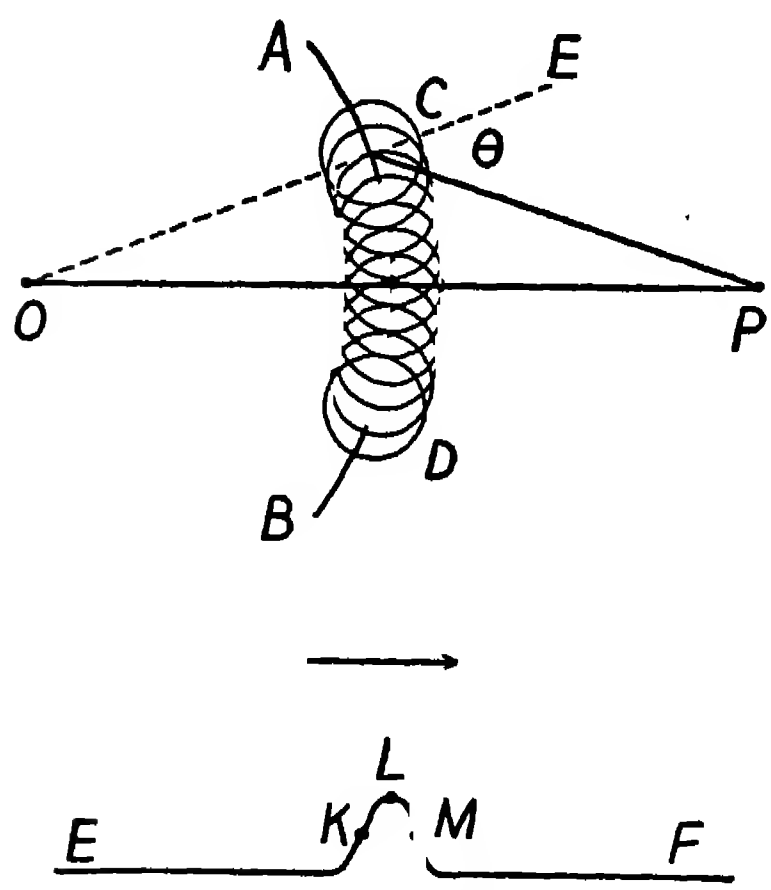


FIG. 188

The difficulty about the back wave is not a serious one from a general point of view. Let the curve EF in the lower part of the figure represent the crest in cross-section. The point L is momentarily at rest, the point K is moving down, and the point M is moving up. Now the downward motion of K and the upward motion of M favor propagation forward and oppose

propagation backward and the result of the state of motion which prevails throughout the entire crest is, as appears both from actual observation and upon applying general equations of wave motion, that the crest moves forward only. There is actually no back wave. But it is by no means evident in what manner this circumstance is to be taken account of in applying Huygens' principle. What Fresnel did was, as explained in chapter 2, to consider each wavelet—for example, the wavelet C—as having an amplitude which is a maximum in the direction CE normal to the wave front, and gradually decreasing as the angle of inclination θ increases, falling to zero when $\theta = \pi/2$ and remaining zero for all backward directions. This procedure, while it works very well, is somewhat arbitrary. We shall learn later of the result to which more rigorous treatments lead.

Now in regard to the difficulty of expecting a feeble prolonged disturbance to pass through any given field point, in place of the well-defined crest which actually passes: The difficulty is of course not with the principle itself but with the naïve manner in which it was applied. If we had a way of taking into account not only the displacement at the peak of the crest, but the displacement velocity and acceleration of all points throughout the crest, we should not encounter this difficulty. It is evidently not permissible to apply Huygens' principle, in the naïve manner above supposed, to a single pulse.*

Fresnel's theory of diffraction does not encounter the last-mentioned difficulty at all. His theory is explicitly based upon a train of waves, and when Huygens' principle and the principle of interference are applied in conjunction to a train of waves this particular difficulty simply does not arise.

Fresnel's manner of applying Huygens' principle, however, leads to an error of $\pi/2$ in the phase with which the resultant disturbance arrives at any given field point. This fact was pointed out at the end of section 12 of chapter 2. When Huygens' principle is rigorously formulated this error of phase is corrected. Moreover, the rigorous formulations indicate that when Huygens' principle is applied after the manner of Fresnel, the wavelets should each be considered to start in a phase advanced by one-quarter of a period over that which prevails upon the wave front.

It has been shown by A. Gouy† and others that whenever a wave passes through a focus an advance of phase of one half-period occurs. This ad-

* In discussing the action of a grating and of a prism on a single pulse, Huygens' principle was applied in a simple manner, but the supposed situation was then different. The interval of arrival of the disturbances was considered in the plane in which a lens would bring the primary wave to a focus. If, in our present case, we consider a suitable lens placed in such a position between the primary wave and the field point that the primary wave would be brought to a focus at the field point, the difficulty would disappear.

† *Compt. rend.*, **110,** 1251, 1890.

vance does not occur discontinuously, but occurs in an interval extending about a wave-length each way along the direction of propagation. Now in applying Huygens' principle to a primary wave front this wave front is of course not passing through a focus, but the secondary wavelets are, so to speak, originating from a focus. Accordingly, it is not unreasonable to assign to wavelets which are, so to say, originating from a focus, one-half of the total advance in phase which occurs when a wave passes through a focus.

The application of the principle of interference.—There is also a certain lack of rigor in the application of the principle of interference which Fresnel's theory makes. In applying this principle no account is taken of the fact that disturbances reaching a given field point from different parts of the wave front are inclined to each other in direction. However, when the wave-length is extremely short in comparison to the distance of the field point from the diffracting system, as it always is in phenomena involving the diffraction of light, the failure of the theory to take account of this relative inclination is of no practical importance, because the magnitude of the final result is determined by disturbances which make only very small inclinations with each other. It is to be remembered in this connection that Fresnel's theory introduces an undetermined obliquity factor, and the same circumstances which make it permissible to leave this factor undetermined also make it permissible to leave out of account the relative inclinations of the disturbances. And this applies equally whether we consider light-waves as longitudinal, as Fresnel did when he developed his theory, or whether we consider them as transverse, as we now feel confident that they are.

The theory is scalar.—As pointed out upon previous occasions—for example, in chapter 1, section 5—Fresnel's theory applies equally well to longitudinal and to transverse waves. This is a point of strength in so far as it permits the theory to be applied, under proper restrictions, to waves of either type. But it is of course a point of weakness in so far as the theory leads to no way at all of arriving at conclusions regarding the state of polarization of diffracted transverse waves. The theory, as such, fails to assign a definite direction in space to the resultant disturbance, with either longitudinal or transverse waves, and it is hence a *scalar* theory, whereas a vector theory is in last analysis to be desired. With longitudinal waves the direction of the diffracted disturbance may be assigned, however, within fairly close limits from general considerations.

With transverse waves, general considerations, when invoked, must be applied with extreme caution. Consider a beam of light incident perpendicularly upon a diffracting screen bounded by a straight edge. Let the incident light be plane polarized with the vibration plane making an angle of, for example, 45° with the edge. At a point within the geometrical shadow the

vibration must, because the waves are transverse, lie in the plane perpendicular to the line joining the field point to the edge. One is now tempted to conclude that the diffracted light will be plane polarized with the vibration in that plane which passes through the direction of vibration of the primary light and the line joining the field point to the diffracting edge. At small angles of diffraction this conclusion would be borne out by observation to within the errors of measurement. But what does this conclusion mean? It means that at small angles of diffraction the diffracted light will have practically the same state of polarization as the incident light, or, at small angles, no marked effects of polarization arising from diffraction are to be expected. And, actually, no marked effects occur. But this is as far as general considerations lead. At large angles of diffraction the material and form of the edge must be considered, for reasons previously explained. Marked effects of diffraction upon the state of polarization occur, but general considerations do not allow us to draw any conclusions regarding the nature of these effects.

Summary and conclusion.—Fresnel's theory suffers from the following limitations and shortcomings: It does not apply at large angles of diffraction; nor close to the diffracting edge; nor to apertures or obstacles which have dimensions of only a few wave-lengths; it fails to give the phase correctly; the theory is scalar and thus fails to take any account of polarization effects—and such subsidiary considerations which may be invoked to supplement the theory lead only to the conclusion that the polarizing effects at small angles of diffraction will themselves be small. As over against these shortcomings, Fresnel's theory, while lacking rigor, is based upon reasonable principles and explains satisfactorily the great body of ordinary diffraction phenomena. It, moreover, explains these phenomena in a manner which is as simple as can be hoped for, considering the rather difficult nature of the general problem, and it explains them in a manner which permits the formation of a concrete physical picture. These latter factors secure for it the feature of being still the most manageable treatment which we have, and hence also the most powerful treatment in the solution of actual diffraction problems.

We pass on now to consideration of what comparatively recent and more rigorous treatments achieve.

26. The Dynamical Theory of Diffraction; General.—The dynamical theory of diffraction is not a *single* theory. It is the aggregate of numerous dynamical treatments of which the two principal ones—the one of Kirchhoff and the one of Sommerfeld—are broad in their point of view. The remaining ones treat special problems, but all represent a great improvement over Fresnel's so-called "kinematical theory" in respect to the rigor with

which underlying principles are formulated and conclusions are drawn. These treatments, taken collectively, place the subject of diffraction upon a much more satisfactory basis from a general or let us say philosophical standpoint; however, most of them involve idealizations, and hence also have their limitations—and they are very elaborate mathematically. The treatment of Sommerfeld is so difficult that it has thus far been applied to only two specific cases, namely, to the single straight edge and to the single slit.

The first noteworthy step toward creating a dynamical theory of diffraction was taken in the year 1849 by G. G. Stokes.* The first part of Stokes's treatment has for its object the deduction of the law according to which the amplitude of the Huygens' secondary wavelet must be considered to vary with obliquity. Stokes arrived at the conclusion that the amplitude of the secondary wavelet is proportional to $(1+\cos\theta)$, where θ is the angle of inclination with the normal to the primary wave front. That is, according to Stokes, the amplitude attains the value zero when, and only when, $\theta=\pi$, or only in the straight-backward direction. The amplitude does not fall to zero at $\theta=\pi/2$ and remain zero for all greater values of θ, as Fresnel's theory supposes. Stokes's treatment, moreover, corrects the error of phase which the Fresnel theory makes.

It should be mentioned that the first part of Stokes's treatment is a special case of the more general treatment given by Kirchhoff some thirty-five years later.

The second part of Stokes's treatment concerns itself with questions of the polarization of diffracted light. But in Stokes's theory, as in Fresnel's, it is assumed that the effect of the material of the diffracting screen upon the passing wave may be neglected. As a result of this circumstance the theory fails in regard to the polarization effects which it predicts; this fact was revealed when various investigators subsequently put the theory to test by examining the polarization effects produced by gratings when diffracting light at large angles. Stokes himself later recognized that the basis of his theory is insufficient to allow conclusions regarding polarization effects to be drawn.

Stokes's theory played an important rôle in the discussions of the day upon the question as to which is the vibration plane in a beam of plane polarized light, but the evidence adduced by the theory was eventually adjudged to be inadmissible.†

As for the special problems which the dynamical theory of diffraction

* *Trans. Camb. Phil. Soc.*, **9**, 1, or Stokes, *Math. and Phys. Papers*, **2**, 243.

† For further information and references see Winkelmann, *op. cit.* (1st ed., 1894), **2**, Part I, 840, and *ibid.* (2d ed., 1906), **6**, 1111.

comprises: The treatment of Rayleigh scattering upon the electromagnetic theory falls into this group. Other problems are, for example, the diffraction by a cylinder when taking into account the optical constants of the material, and the corresponding problem for the sphere. Numerous problems of this general type are included.*

27. The Theory of Kirchhoff.—An abstract of this theory will be presented, together with pertinent comments.

This theory, like that of Fresnel and that of Stokes, assumes that the light incident upon the diffracting screen is merely stopped and that the remainder of the wave passes unaltered; the theory is consequently subject to the previously mentioned limitations which this assumption entails. Kirchhoff's theory is also scalar. The important contribution of this theory is that it leads to a more rigorous formulation of Huygens' principle than does Fresnel's theory. In Kirchhoff's theory, Huygens' principle is deduced from the general differential equation of wave motion by making suitable transformations. The treatment is *not* an electromagnetic one.

In the footnote will be found references to Kirchhoff's original paper and to more recent and improved presentations.†

Our comments will be based upon the presentation by Drude and following the notation he uses.

The first part of Kirchhoff's theory is concerned with undisturbed wave propagation; there is supposedly no diffracting screen present. Let s be the instantaneous wave disturbance at any point P in space; s must satisfy the general differential equation of wave motion:

$$\frac{\partial^2 s}{\partial t^2} = V^2 \left(\frac{\partial^2 s}{\partial x^2} + \frac{\partial^2 s}{\partial y^2} + \frac{\partial^2 s}{\partial z^2}\right) \equiv V^2 \Delta s \,, \tag{32}$$

where V is the velocity of waves in the given medium; t is the time; and x, y, and z are the co-ordinates of P in a rectangular system. It is necessary that the value of s be known at all points in space at all instants of time. This value will be known if, by way of example, Q (Fig. 189) be a point source of a simple harmonic disturbance, which, at the point P, at the distance r_1 from the source, yields the disturbance:

$$s = \frac{A}{r_1} \cos 2\pi \left(\frac{t}{T} - \frac{r_1}{\lambda}\right) = \frac{A}{r_1} \cos \frac{2\pi}{T} \left(t - \frac{r_1}{V}\right) \,, \tag{33}$$

* See *Encyk. d. Math. Wiss.*, **5**, Part III, 488, and Geiger and Scheel, *op. cit.*, **20**, 263.

† G. Kirchhoff, *Wied. Ann.*, **18**, 663, 1883; Drude, *Theory of Optics*, trans. Mann and Millikan (1920), p. 179; *Handb. d. exp. Phys.*, **18**, 296; *Encyk. d. Math. Wiss.*, **5**, Part III, 418; and, for general critical comments, Geiger and Scheel, *op. cit.*, **21**, 906.

where A is the amplitude at unit distance from the source, T is the period, λ is the wave-length, and V is as before the velocity of the waves—and $TV=\lambda$. However, in the general case, which is the one treated in the theory at the outset, the value of s need not be as given in equation (33); the original disturbance need not be periodic or originate from a point source—but it must be everywhere known.

Denote the value of s at the particular field point P_0 by s_0; since s_0 is merely the value of s at a particular point, the value of s_0 will of course also be known. The first problem which the Kirchhoff theory solves is to express s_0, not directly from whatever mathematical conditions may be given, but in the form of an integration of s and its derivatives over an arbitrary closed surface S, which surrounds P_0 and excludes the source of the disturbance. Letting $\partial/\partial r$ and $\partial/\partial n$ represent respectively differentiation along the radius vector r indicated in the figure, and along the normal to the surface, n, the following result is arrived at after numerous transformations involving especially Green's theorem:

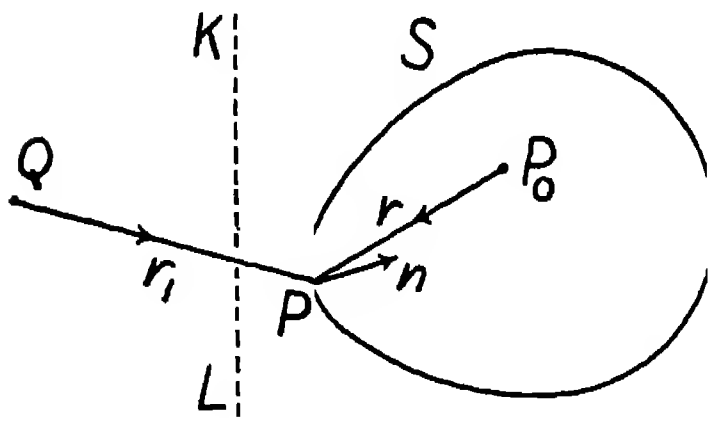


FIG. 189

$$(34)\qquad s_0=\frac{1}{4\pi}\int\int\left[\frac{\partial\dfrac{s\left(t-\dfrac{r}{V}\right)}{r}}{\partial r}\cos(nr)-\frac{1}{r}\frac{\partial s\left(t-\dfrac{r}{V}\right)}{\partial n}\right]dS\,.$$

Here $s[t-(r/V)]$ means: s function of $[t-(r/V)]$. That is to say, the value of s_0 at P_0 which the equation gives is that value which obtains at the instant t, and the value of s which enters in determining s_0 at the instant t is that value of s which obtains at P at a time r/V seconds earlier, because r/V is the time which it takes for a disturbance to travel from P to P_0. Accordingly, the value of the time which must be inserted in whatever the expression for s may be is $[t-(r/V)]$, and not t. Moreover, it is not the value of s itself which is called for, but the value of the derivative of s/r along r and the derivative of s along n.

The basic concept of Huygens' principle is easily recognized in equation (34). According to this equation, the value of the disturbance s_0, at any point P_0, may be regarded as arising from the disturbances sent to P_0 from any surface S across which the primary disturbance must pass in order to reach P.

Let us return now to the case in which the primary disturbance originates

from a point source and is simple harmonic. The value of s will then be given by equation (33). By substituting this value in equation (34)—for details see Drude—this equation becomes:

$$s_0 = \frac{A}{2\lambda} \int\int \frac{1}{rr_1} \sin 2\pi \left(\frac{t}{T} - \frac{r+r_1}{\lambda}\right) [\cos(nr) - \cos(nr_1)] dS . \tag{35}$$

Comparison of this equation with the corresponding one of Fresnel's theory (eq. [13] of chap. 2) is now in order. When equation (13) of chapter 2 is written in integral form and the notation is adapted to our present notation, which follows Drude, it becomes:

$$s_0 = A \int\int \frac{k}{rr_1} \cos 2\pi \left(\frac{t}{T} - \frac{r+r_1}{\lambda}\right) dS . \tag{36}$$

Hence the inclination factor k, which Fresnel left undetermined, has the value

$$k = \frac{1}{2\lambda} [\cos(nr) - \cos(nr_1)] . \tag{37}$$

Moreover, since equation (35) has the sine function where equation (36) has the cosine function, the error of phase of $\pi/2$ made by the Fresnel theory is corrected in the Kirchhoff theory.

We are now ready to consider the portion of Kirchhoff's theory which applies to diffraction proper. When a diffracting screen is erected at, let us say, the position of the plane KL of Figure 189, the value of the disturbance s at every point of the closed surface S will no longer be known. However, the closed surface S is an arbitrary one and may be taken in such a way that the plane KL forms a part of the surface. But by taking this step alone the difficulty is not removed. Therefore, in a sense Kirchhoff's theory requires the answer to the problem to be known before the problem can be solved. The following scheme of circumventing the difficulty in question is now resorted to: The arbitrary surface is considered to coincide with the plane KL to infinity, and then to surround the field point, at infinity. For a surface element at infinity the radii vectores r and r_1 are parallel, and hence the angles (nr) and (nr_1) are equal, therefore $[\cos(nr) - \cos(nr_1)] = 0$, and hence the contribution from the portion of the surface which lies at infinity is zero. Thus the integration needs to be extended only over the plane KL to infinity.

When a diffracting screen is erected, it is assumed that s and $\partial s/\partial n$ both vanish at all points on the side of the screen away from the source, and that

for all points which lie in the apertures of the screen the values of s and $\partial s/\partial n$ are everywhere unaffected by the presence of the screen. This assumption, while reasonable, may, from the point of view of rigor, be objected to, principally on the following grounds: First, the theory, as originally deduced, requires that s and $\partial s/\partial n$ be everywhere continuous, whereas the assumption in question introduces discontinuities at the edge of the diffracting screen. Second, the assumption neglects the effect of the material of the screen upon the passing wave and hence puts Kirchhoff's theory in this regard into the same category with Fresnel's theory.

In view of the foregoing considerations, and especially in view of the marked similarity between equations (35) and (36), it is obvious that Kirchhoff's theory becomes in practice very much like Fresnel's. But we must always bear in mind that Kirchhoff's theory presents a great advance over Fresnel's in regard to the rigor with which Huygens' principle is formulated.

28. Remarks concerning the Theory of Sommerfeld.—This theory is so difficult that we shall not attempt to submit even an abstract of it. However, a few brief comments will be in place. References to both the original and later presentations are given in the footnote.*

From the point of view of our present comprehension of the nature of light, the treatment of any problem in diffraction is considered as leaving nothing to be desired when it is based upon the electromagnetic theory and furthermore takes into account the complete form of the diffracting object and the optical properties of the material of which this is made, without unrealizable idealization, and is mathematically rigorous. The theory of Sommerfeld almost satisfies these desiderata, but it makes idealizations in regard to the form and optical properties of the diffracting screen which cannot be quite realized, nor even practically realized from all points of view.

The diffracting screen is assumed to be infinitely conducting, that is, perfectly reflecting, and, in addition, either infinitely thin, or otherwise, having the form of a wedge of given angle, the wedge coming to a theoretically perfect edge. Granted the possibility of realizing such screens, the theory leaves nothing to be desired.

At the outset a screen of any contour is considered, with the source, a point source, occupying any position. Subsequently the problem is restricted to a screen bounded by a single straight edge with the source removed to infinity in any direction perpendicular to the edge and the incident light being plane polarized. There are then two cases considered: (*a*) electric vector parallel to the diffracting edge and (*b*) electric vector perpendicular to the diffracting edge.

* Sommerfeld, *Math. Ann.*, **47**, 317, 1896; Riemann-Webers, *Diff. Gl. d. Phys.*, **2**, 433, 1927 (by Sommerfeld); *Encyk. d. Math. Wiss.*, **5**, Part III, 488 (by P. S. Epstein); Geiger and Scheel, *Handb. d. Phys.* **20**, 266 (by G. Wolfsohn).

When the formulae finally obtained are applied for small angles of diffraction, they bear out the results to which the less rigorous treatments of Fresnel and of Kirchhoff lead; the error of phase which the Fresnel theory makes is, however, corrected.

The theory of Sommerfeld reveals clearly that a wave will appear to come from the diffracting edge, after the manner described in section 11. (It should be mentioned in this connection that the Kirchhoff theory also reveals this fact, as was first pointed out by E. Maey.)

The predictions which the Sommerfeld theory makes for large angles of diffraction have been compared with experiment. The result of the comparison will be spoken of in the next section.

The case of the single slit has been treated by K. Schwarzschild.*

29. Diffraction at Large Angles.—In order to have sufficient intensity to permit measurements upon the relatively minute amount of light which is diffracted to large angles, the incident beam is focused upon the diffracting edge after the manner illustrated in Figure 190. Light from a carbon arc A is focused upon a slit S having a width of perhaps 1 mm, and the light from this slit is in turn *focused* upon the edge of the diffracting screen D. The experimental conditions which prevail thus differ from the usual ones and likewise from those hypothecated in Sommerfeld's treatment, which are similar to the usual ones. However, the fact that the light is focused upon the diffracting edge appears not to be a disturbing factor; it is found that the amount of light which is diffracted to any given large angle does not depend critically upon the exactness of the focal adjustment. Much more important when making comparison between theory and experiment is that the requirement that the edge be perfectly sharp and the material perfectly reflecting cannot be fulfilled. It is presumably on this account that observation does not bear out the predictions of theory very closely.

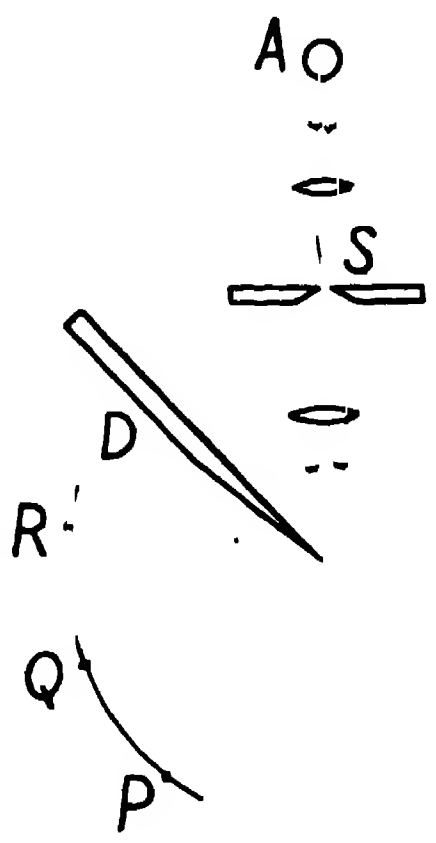


FIG. 190

Observations and measurements upon the light diffracted to large angles were made by A. Gouy, by W. Wien, and by E. Maey before the theory of Sommerfeld had been developed. More recently A. Kalaschnikow, F. Jentsch, and others have carried out measurements of the same type.

For obvious reasons of convenience, attention has been directed primarily to the light which is inflected rather than to that which is deflected.

* *Math. Ann.*, **55**, 177, 1902. For a summarized account see Geiger and Scheel, *op. cit.*, **20**, 281.

When the incident light is plane polarized with the electric vector parallel to the diffracting edge, theory and experiment agree fairly well. However, the amount of light which is inflected falls off somewhat more rapidly with increasing angle of inflection than theory calls for.

When the electric vector is perpendicular to the diffracting edge, the amount of light inflected should, according to theory, be several times greater at a given angle than the amount inflected at the same angle when the electric vector is parallel to the edge. By way of comparison, experiment reveals that when the electric vector is perpendicular, there is actually more light inflected than when the electric vector is parallel, but the factor of excess is decidedly smaller than that which theory calls for.

For further information and detailed references see footnote.*

Kalaschnikow has described experiments which demonstrate in a striking and entirely objective manner that the diffracted light appears to come from the edge itself. Referring again to Figure 190, he mounted a photographic plate in the plane of the figure, that is, perpendicular to the diffracting edge, and erected ordinary sewing needles at, for example, the points P, Q, and R. Upon developing the plate he found that each needle had cast a shadow in such a direction as to show that the light had come from the diffracting edge; the shadows all radiate out from the edge.

We pass on now to the discussion of another case which, like the one we have been considering, falls outside of the realm of the ordinary diffraction phenomena of either the Fresnel or the Fraunhofer class.

30. Observation Close to the Diffracting Edge.—The kind of material of which the diffracting edge is made influences the distribution of light in the region close behind the edge for the reason explained in section 25. Such an influence would exist even if the edge were a theoretically perfect one. But in practice the edge is never perfect and its lack of perfection introduces additional factors, expecially one of reflection from the edge to the portion of the pattern which lies outside of the geometrical shadow.

The light which is reflected by the edge diverges as from a virtual image located close to the edge, and since it diverges from a point so situated, this light decreases rapidly in intensity as one recedes from the edge.

When the edge is blunt, or has a large radius of curvature, more light is reflected than when the edge is sharp, or has a small radius of curvature. Hence also, when the edge is blunt the reflected light maintains sufficient intensity to affect the pattern appreciably, for a distance farther back than when the edge is sharp. Moreover, a highly polished surface reflects better than one which is mat, but since even a mat surface reflects well when light

* Geiger and Scheel, *op. cit.*, **20**, 274. Here π is used to denote "electric vector parallel to the diffracting edge" and σ is used to denote "electric vector perpendicular [*senkrecht*]."

grazes it, as it does in the case with which we are concerned, the question of whether the edge is polished or mat is of less importance than whether it is blunt or sharp.

To understand the effect of reflection, let us first consider a large plane mirror erected vertically and employed as a Lloyd's single mirror for producing interference fringes. Let the beam of light be horizontal and, accordingly, the fringes vertical. A pure interference pattern without diffraction will be formed by this mirror. Now let the mirror gradually take on cylindrical curvature in such a way that the elements of the cylinder are vertical. A diffraction pattern will presently superpose itself upon the interference pattern. Moreover, as the mirror becomes more and more curved the reflected light will become more and more divergent and hence fainter and fainter. Consequently, if we are at a distance of some decimeters from the "diffracting edge," the reflected light will soon be too faint to modify the diffraction pattern noticeably and we shall find ourselves confronted with a practically pure diffraction pattern. But by approaching the "edge" more closely, a region may be reached where the interference pattern still has appreciable intensity and modifies the diffraction pattern. That is, we should always expect to find reflection modifying the diffraction pattern in a region sufficiently close to the edge.

The problem which we have before us has been investigated both experimentally and theoretically by N. Basu* and by T. K. Chinmayan.†

The effects which come into question may conveniently be observed by using as a mirror a strip of plate glass which is perhaps two inches wide and a yard long. It is preferable in practice to mount the mirror horizontally instead of vertically. The mirror may then be laid upon knife edges and be bent as desired by loading it appropriately. It is not necessary that the glass be silvered.

31. Diffraction and the Quantum Theory.—By way of concluding our consideration of the diffraction of light, a few words will be said regarding the relation which exists between diffraction theory and the quantum theory.

The fundamental concepts of diffraction theory proper were all originally propounded in the belief that light is purely a wave phenomenon, whereas we now believe that light has a dual nature of waves and of corpuscles. The fact that light has this dual nature has become evident since the advent of the quantum theory in the year 1900. Accordingly, the question arises as to whether the quantum theory calls for any modifications of the fundamental concepts of diffraction theory as originally propounded, and as submitted in this book. These concepts in their most advanced state of development

* *Phil. Mag.*, **35**, 79, 1918. † *Ibid.*, **37**, 9, 1919.

are based upon the application of the electromagnetic wave equations of Maxwell to particular problems. And these equations, before the advent of the quantum theory, represented our ultimate knowledge. We have now, however, in certain respects gone beyond Maxwell's equations, and it has developed that Maxwell's wave equations represent the correct average result whenever we are concerned with a large number of quantum processes considered collectively, but that in the individual processes of emission and absorption, light has certain particle-like properties, which are revealed in experiments of various types such as, for example, those of the photo-electric effect and of the Compton effect.

The wave nature and the corpuscular nature of radiation stand today completely reconciled to each other by the quantum theory; and, incidentally, it has developed that matter has the same dual nature which light has—and with matter also, the two aspects of its nature stand reconciled to each other.

The purpose of diffraction theory is to account for the distribution of intensity which obtains under each of certain sets of given conditions. Now the question of the intensity of the electromagnetic radiation which obtains at any given field point or in any given direction is a question of the average result of a large number of individual quantum processes, and this average result is, as above stated, always expressible by Maxwell's equations. Hence, when any problem involving the diffraction of light is solved in terms of the electromagnetic theory, we may today, just as in the day when the electromagnetic theory represented our ultimate knowledge, consider the problem as solved in a completely satisfactory manner.

CHAPTER 8

THE DIFFRACTION OF X-RAYS

Before entering upon our subject proper, let us take a brief look backward historically.

1. Roentgen's Discovery.—On November 8, 1895, Wilhelm Conrad Roentgen, professor of physics, then at the University of Wurzburg, Bavaria, was experimenting with a cathode-ray tube actuated by a Ruhmkorff induction coil when he discovered a startling phenomenon. Precisely what Roentgen first observed appears to be in some doubt, but most probably he was at the time working in a well-darkened room with a cathode-ray tube which he had covered all over with opaque black paper, when, to his great astonishment, he observed that a barium platinocyanide screen which was lying near the tube showed a distinct fluorescence. He soon convinced himself that the cause of the fluorescence originated from the cathode-ray tube, and not from the Ruhmkorff coil or other parts of the electrical circuit.*

Roentgen followed up his first observation, whatever it may have been, by seven or eight weeks of intensive investigation, and then published his findings. He had ascertained, in the first place, that what caused the fluorescence of the screen was of the nature of a *radiation* of which the source was located where the cathode rays struck the wall of the glass tube, inside, or struck the anode of the tube. And he had further ascertained that the rays in question have a great power for penetrating all materials, but that they penetrate dense materials, in particular the metals, much less readily than they penetrate light materials. To demonstrate this he made shadow photographs of brass balance weights through their wooden containing box. Thereupon he photographed the bones of the hand through the flesh.

Roentgen also investigated various electrical effects produced by the rays, studying in particular the power which the rays have of ionizing gases and thus causing electrically charged bodies to lose their charge.

Finally, he attempted to refract the rays by prisms made of various materials but failed to find any refraction which he could measure, and he also failed to find any specular reflection. But he discovered that diffuse reflec-

* For further information regarding the probable circumstances of Roentgen's discovery see the book entitled *Wilhelm Conrad Roentgen u. die Geschichte der Roentgenstrahlen* (Berlin: J. Springer, 1931), by Otto Glasser, in particular pp. 3–12 and 36.

tion or scattering occurs within the body of material which the rays penetrate—similar to the scattering of light from within the body of a turbid medium.

He himself gave to the new rays the name "X-rays."

Roentgen's discovery became known to the world at large during the first week in January, 1896, and caused unparalleled excitement both scientifically and popularly.

2. The Nature of X-rays.—After Roentgen's discovery the question of course arose immediately as to what the fundamental nature of X-rays might be, and, in particular, as to whether they were material particles of some kind or were due to waves. In the year 1897 G. Stokes,* and in the following year J. J. Thomson,† advanced the hypothesis that X-rays are a form of electromagnetic-wave disturbance originating where the fast-traveling cathode particles, or electrons, strike the wall of the tube or the anode, and further that the X-ray disturbances caused by this impact have the form of extremely short separate pulses rather than of wave trains. According to this view, one pulse supposedly originates each time a cathode particle has its velocity suddenly altered by striking the wall of the tube or the anode, the particle being there either stopped or caused to rebound. Since in this operation a large acceleration of an electric charge would be involved for a very brief period of time, an intense and extremely short pulse should be sent out, in accord with well-established electromagnetic theory. The pulses were supposedly much shorter than any of the wave-lengths associated with light.

If X-rays have an electromagnetic wave nature one might expect that they should exhibit reflection, refraction, and diffraction. But even in the early days it was realized that if the waves or pulses are extremely short, this fact might cause the expected effects to become manifest only under rather extraordinary conditions, and such has since proved to be true. Numerous attempts were made during the first fifteen years following Roentgen's discovery to reflect, refract, and diffract X-rays, but they failed completely—except for one experiment of which we shall have occasion to speak further in section 23. In this the investigators claimed to have found a very slight positive effect, but others for good reason claimed that the observed effect might be entirely of photographic origin.

It was also found that X-rays are not deviated by a magnetic field; hence they could not in any event be due to charged particles.

The view that X-rays have an electromagnetic wave nature received

* *Proc. Manchester Lit. and Phil. Soc.*, **41**, No. 15, 1, 1897.

† *Phil. Mag.*, **45**, 172, 1898.

further support when in the year 1906 C. G. Barkla* showed that X-rays can be polarized.

Barkla demonstrated the polarizability of X-rays by a double scattering experiment: Referring to Figure 191, let i be a beam of X-rays which originates at the anode of a cathode-ray tube and is incident upon a first body, A, by which it is scattered. A portion of the radiation scattered by A is incident upon the second scattering body, B. (The body B is of course shielded from the direct radiation coming from the tube.) The original beam i is unpolarized, but if the disturbance in this is electromagnetic and is consequently transverse to the direction of propagation, then, supposing that the scattering is due to displacements of the electrons within the scatterer, the electric vector in the beam scattered by A toward B should be entirely in the plane perpendicular to the plane of our figure, provided no double scattering takes place within the body A. The beam from A incident upon B should then be completely polarized. If complete polarization exists, this will be revealed by the fact that the polarized radiation incident upon B from A will be scattered by B with maximum intensity in the directions c and d, and with zero intensity in the directions from B perpendicular to the plane of the figure. If the polarization is incomplete, there will merely be a difference in intensity exhibited. In Barkla's work only a difference was observed, because double scattering occurred, and because, in order to secure sufficient intensity to carry out the experiment, it was necessary to use beams subtending quite large solid angles, and thus the ideal geometry of the scheme was not strictly maintained. Much later experiments by Compton and Hagenow† showed that under proper conditions the polarization is complete.

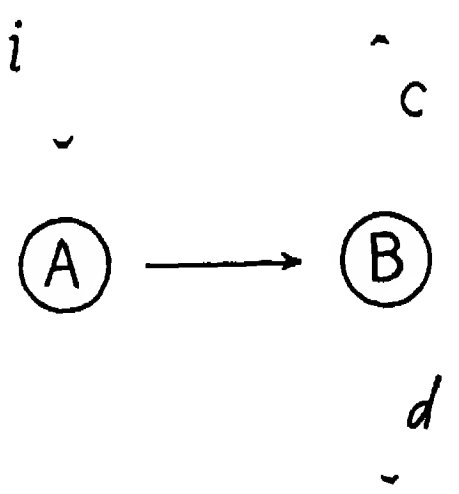


FIG. 191

Our understanding of the nature of X-rays has by the present day advanced so far that we may say with positiveness that X-rays have an electromagnetic wave nature; but the former isolated pulse hypothesis must be revised. It was discovered, in the year 1912, that X-rays may be diffracted by *crystals*, and through the discovery of this diffraction, which we shall consider in detail later, it became possible to obtain X-ray "spectra." These spectra consist, like optical spectra, of discrete spectrum lines superposed upon a continuous background. Now the radiation representing each spectrum line is evidently homogeneous and hence must arise from a long train

* *Proc. Roy. Soc.*, **77**, 247, 1906.

† *JOSA and RSI*, **8**, 487, 1924.

of waves. The continuous background, on the other hand, presents a certain analogy to the continuous spectrum of the optical region, or, to "white light," and if the analogy were complete, the continuous background could be considered as arising from isolated pulses, but the analogy is not complete. The continuous optical spectrum fades gradually at both ends without definite limit at either end, and we may consider this spectrum as due either to trains of waves of all wave-lengths or to isolated pulses, as explained in chapter 7, sections 22–24. For this spectrum the hypothesis of isolated pulses probably corresponds somewhat more nearly to physical reality. The continuous X-ray spectrum, on the other hand, while fading gradually at the end of long wave-lengths, ends completely at a well-marked limit at the end of short wave-lengths. Such a complete and well-marked termination could not occur if this continuous spectrum arose from isolated pulses, and it is hence nearer the truth to think of the continuous X-ray spectrum, as well as the spectrum lines, as arising from long trains of waves.

The position of the short wave-length limit of the continuous spectrum is determined by the fundamental quantum relation:

$$E = h\nu = \frac{hc}{\lambda}, \tag{1}$$

where E denotes the kinetic energy of the electrons bombarding the anode of the cathode-ray tube, as determined by the voltage applied to the tube; h is Planck's constant; ν is the frequency and λ the wave-length corresponding to the position in the spectrum at which the limit occurs; and c is the velocity of electromagnetic waves in vacuum, that is, the velocity of light.

The nature of a given X-ray emission spectrum is determined by the kind of material of which the anode or "target" of the cathode-ray tube, or "X-ray tube," is made. Each type of atom—for example, Fe, Cu, Ag, etc.—which may be present in the material, emits its own characteristic spectrum of lines of given position and relative intensities regardless of whether the type of atom in question is present in its elemental state or in chemical combination. That is to say, X-ray spectra are *atomic* spectra.

The spectrum of a given chemical element may also be obtained by placing the element itself or a compound containing the element upon the target of an X-ray tube where it will be bombarded by the cathode particles.

Each chemical element which is present within or upon the target may give rise to two series of spectrum lines which are designated by the letters K and L, the K series being the one of shorter wave-length and the L series the one of greater wave-length. Additional series in so far as found will be spoken of later. The several series occupy entirely distinct spectral positions, that is, there is no overlapping.

Within each series the spectrum lines appear in groups which are usually well separated. The group of greatest wave-length is denoted by the letter α, the next group by β, etc. The individual spectrum lines of each group are given numerical subscripts. Accordingly, X-ray spectrum lines receive such designations as Fe $L\alpha_1$, Cu $L\beta_3$, Ag $L\gamma_2$. The numerical subscripts within each group are assigned in the order of intensity of the lines, beginning with the most intense line of the group as number one.

Corresponding spectrum lines of different elements, for example, Al $K\alpha_1$ and Cu $K\alpha_1$, have frequencies of which the square roots are very nearly proportional to the atomic numbers of the elements. This important relation was discovered by H. G. J. Moseley* and is known as "Moseley's law."

From Moseley's law it follows as a corollary that the spectra of the various elements will show striking features of similarity. In addition, X-ray spectra are much simpler than optical spectra and hence are much more readily systematized.

For the heavier elements, for example, Pt and Au, two additional series of lines have been found which are designated by the letters M and N. These series occur, in the order given, at greater wave-lengths than the L series.

The K series of spectrum lines is the simplest one. It has a single definite limit on the side of short wave-lengths and none of the lines of the series appears until the energy of the bombarding cathode particles attains the value determined by the quantum relation, equation (1), in which ν, or λ must be given the value which corresponds to the series limit. Then the entire series appear at one time. The process which causes the lines of the series to appear is supposedly as follows: A cathode particle impinges upon an atom of the target and ejects an electron from the innermost or first electron shell of the atom, the so-called K shell. When an electron from the second shell, or the third, etc., falls to the K shell to replace the electron which was ejected, one of the spectrum lines of the K series is emitted. (The second shell is called the L shell, the third the M shell, etc.) Which particular spectrum line of the K series is emitted depends upon which shell and which "level" of the shell the replacing electron comes from. Since the bombarding cathode particles must have a certain minimum energy before they can eject electrons from the K shell, none of the lines of the K series appears until the energy, or the voltage, attains the proper value. Once the voltage attains this value or exceeds it, there will be an electron ejected from the K shell of each of a multitude of atoms. In some of these atoms the replacing electrons will come from each of the levels of the second, third, etc., shells. Hence the entire series of spectrum lines now appears at once.

The L, M, and N series each supposedly originates in a manner some-

* *Phil. Mag.*, **26**, 1024, 1913, and **27**, 703, 1914.

what similar to that in which the K series originates, but with these other series the conditions become progressively more complicated. Whereas there is only a single K level, there are three L levels and five M levels. Correspondingly, there are three series limits for the L series and five for the M series. For the N series there are probably seven limits, but whether this be true is not definitely known. For further information concerning the supposed mechanism of the emission of X-ray spectrum lines see the reference given in the footnote.*

All X-ray wave-lengths are extremely short when compared to the wave-lengths of the visible spectrum. The visible spectrum, it will be recalled, extends from about 3800 Å to about 7500 Å. In comparison, the X-ray wave-lengths with which we shall be concerned range from about .1 Å to 150 Å. But this is a range of more than ten "octaves," whereas the visible spectrum covers a range of barely one octave.

X-ray wave-lengths are frequently given in terms of the so-called X-unit, which is equal to one-one-thousandth of an Ångstrom unit, or, since the Ångstrom equals 10^{-8} cm, the X-unit equals 10^{-11} cm.

The power of X-rays for penetrating matter increases as the wave-length decreases. X-rays of relatively great wave-length have a low penetrating power, and these are said to be "soft," whereas those of relatively small wave-length have a high penetrating power and are said to be "hard." The X-rays used for medical purposes are chiefly of the latter type.

Even before it was possible to obtain X-ray spectra, C. G. Barkla was able to prove from studies of the "hardness" of X-rays, as determined directly from their power of penetrating matter, that each chemical element gives rise to two kinds of what he termed "characteristic" radiations, which he designated by the letters K and L. What Barkla called K radiation was later analyzed spectroscopically and is identical with what we now call the K series of spectrum lines, and similarly the L radiation was analyzed, and is identical with what we now call the L series of spectrum lines.

While no sharp boundaries occur which serve to classify X-rays into natural groups according to their hardness, it is convenient for purposes of discussion to set up such boundaries more or less arbitrarily. Accordingly, we shall henceforth designate X-rays having wave-lengths of less than 1 Å as hard X-rays, those having wave-lengths between 1 Å and 15 or 20 Å as soft X-rays, and finally, those having wave-lengths greater than some 15 or 20 Å as very soft X-rays.

In X-ray spectroscopy the region of soft X-rays is the most important because the spectral series of so many different chemical elements fall into this region. The spectral regions into which the K, L, and M series of the

* A. Sommerfeld, *Atombau und Spektrallinien* (4th ed.; Braunschweig: Vieweg, 1924), pp. 244 ff.

various elements fall are indicated by means of the graphs shown in Figure 192. Here wave-length is plotted as abscissa, against atomic number, increasing downward, as ordinate. In the region of hard X-rays all of the lines of a given series are comprised within a range of several hundredths or perhaps a few tenths of an Ångstrom unit, and in the region of soft X-rays within several tenths of an Ångstrom or perhaps one or two Ångstroms. For convenience of reference a periodic table of the elements, giving atomic numbers, is given to the left of Figure 192. (This table is somewhat abridged.)

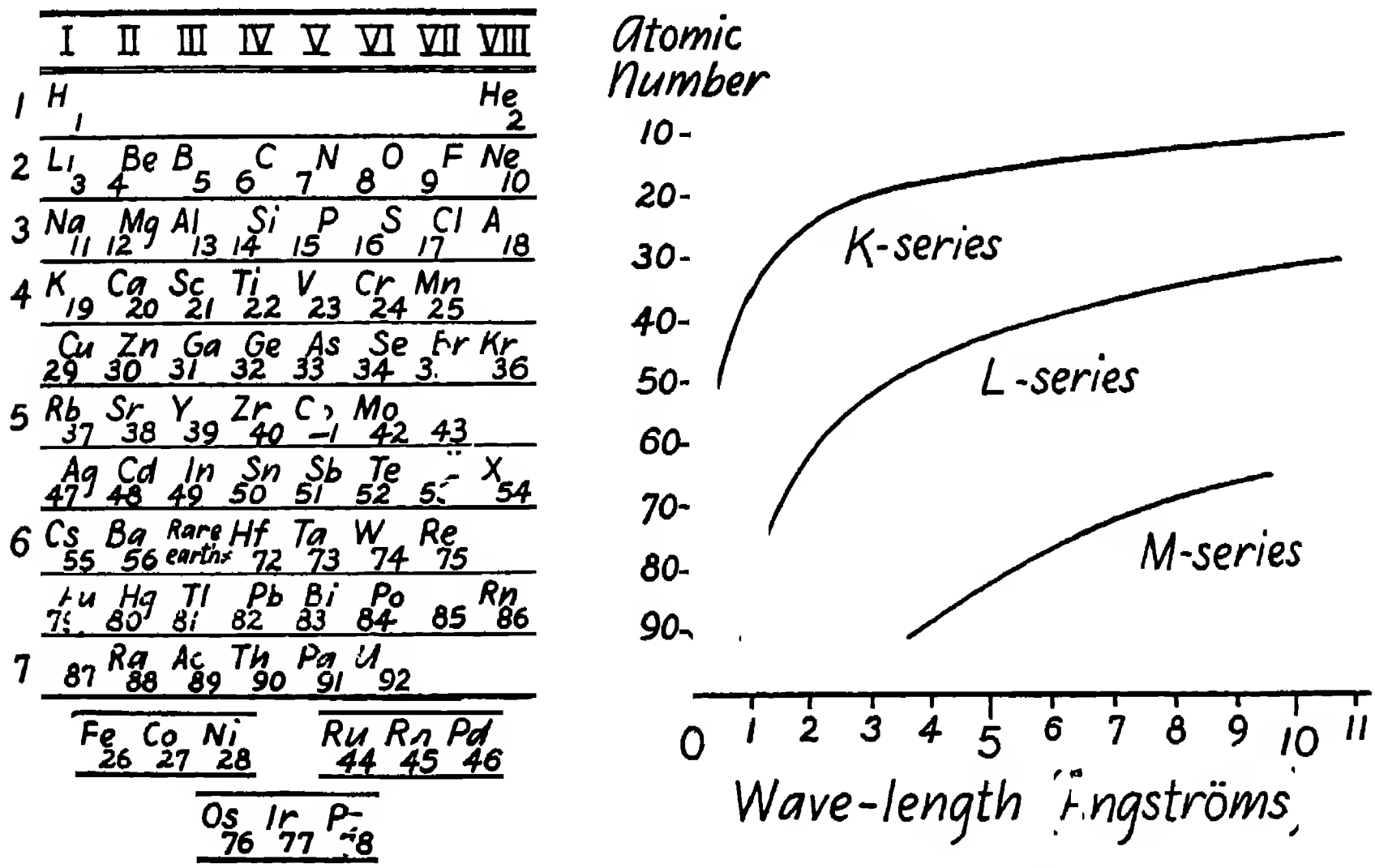


FIG. 192

The *absorption spectra* which result when X-rays pass through matter have also been extensively investigated. Each chemical element produces its own effect characteristic of the atom of which the element is constituted, almost though not entirely regardless of any state of chemical combination which may prevail. That is, X-ray absorption spectra are also in the main *atomic* spectra. For each type of atom the absorption spectrum is closely related to the emission spectrum but the absorption spectrum *does not* consist of the lines of the emission spectrum reversed as is so often the case in optical spectra. X-ray absorption spectra have the following nature: In passing through the spectrum in the direction of increasing wave-length, the amount of absorption, for each element, *increases gradually* until the position of a *limit* of one of the emission series of the element is reached. Here the absorption *very suddenly* falls much lower, producing a so-called

"absorption edge." According to the quantum theory, the energy E required to excite a given emission series must be equal to or greater than $h\nu$, or hc/λ, where ν and λ have the values which obtain at the limit of the series. In other words, the absorption falls much lower at that spectral position at which the energy, or the frequency, of the X-rays becomes too low for the incident waves to excite the spectral series in question in the absorber, as secondary X-rays. Thus, X-ray absorption spectra consist of numerous "absorption edges." Of these, many, for reasons which we shall not enter into, show "fine structure" which has been the object of much careful study.*

We now come to speak of the epoch-making discovery which made our present knowledge of X-ray spectra possible.

3. Laue's Discovery.—In the year 1912 M. von Laue advanced a brilliant idea which, when put to test under his guidance by his students W. Friedrich and P. Knipping, opened a totally new era in our knowledge of X-rays.

The atoms or molecules of a *crystal* are arranged in a regular array in three dimensions after some such fashion as represented in Figures 193 and 194. Here the circles represent the particles; the connecting lines are

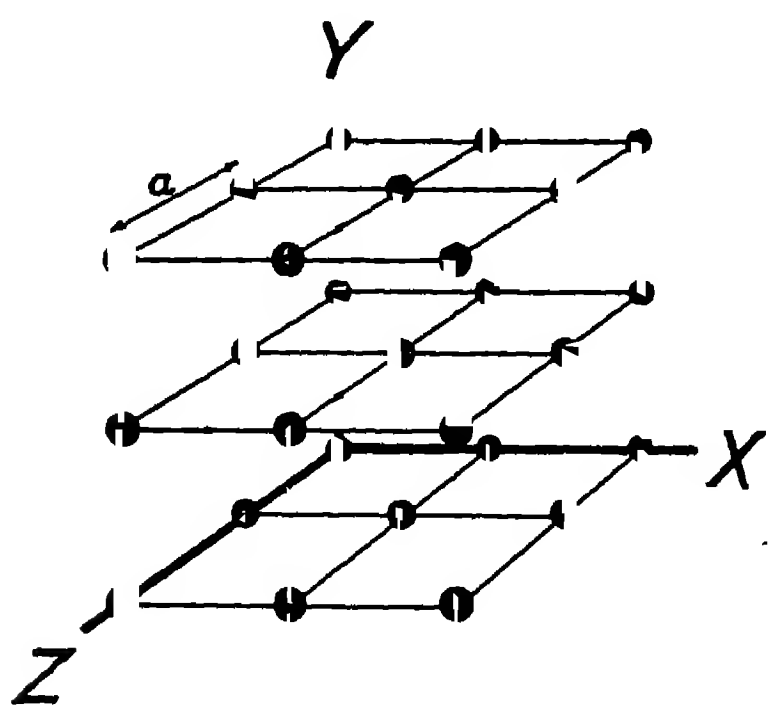


FIG. 193

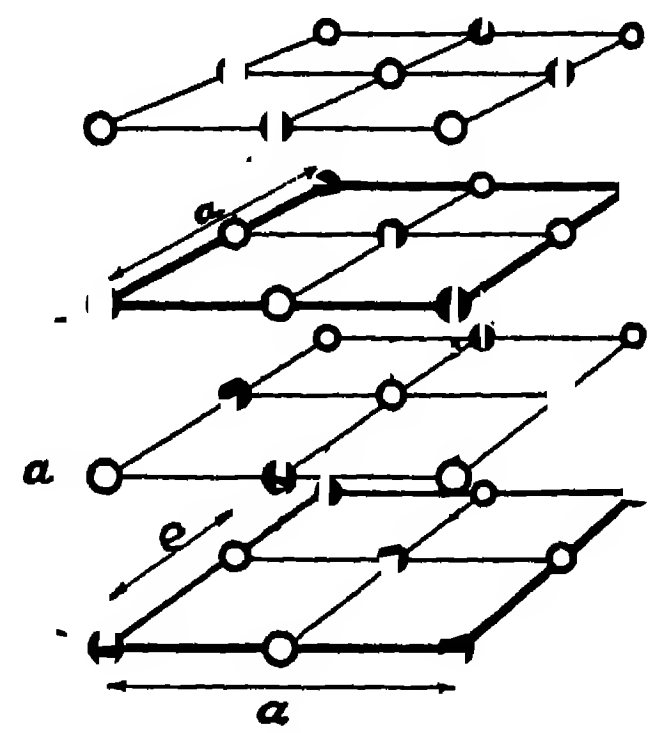


FIG. 194

introduced merely to clarify the drawing. Such an array of particles constitutes what is known as a "space lattice." That the atoms or molecules of a crystal must be arrayed in a space lattice had been believed for a half-century—as a result of studies of the forms which crystals take. And calculation based upon Avogadro's number and the densities of crystals had indicated that the particles of the lattice must be separated by distances of not

* See, e.g., Kievit and Lindsay, *Phys. Rev.*, **36**, 648, 1930; J. D. Hanawalt, *Jour. Franklin Inst.*, **214**, 569, 1932; Coster and Veldkamp, *Zeit. f. Phys.*, **74**, 191, 1932. For the general theory of X-ray absorption spectra see Sommerfeld, *op. cit.*, pp. 288 ff.

more than 5 or 10 Ångstrom units. Now Laue's brilliant idea was as follows: Suppose each atom of any material should act as a center of diffraction or "center of scattering" for X-rays, giving rise to Huygens wavelets, which spread out in all directions about the particle as center. Then, a crystal, in which the particles must be arrayed in a space lattice, may serve to resolve the extremely short waves or pulses of which X-rays are presumably constituted, in a manner somewhat analogous to that in which a diffraction grating resolves the much longer waves of *light*.

Laue developed his idea by applying the principle of interference, and ascertained by mathematical steps involving only ordinary relations of analytic geometry, that the Huygens wavelets should reinforce each other in certain specific directions, and thus give rise to principal maxima of diffraction in these directions. When Friedrich and Knipping set up the experiment, the effects predicted by the theory were found. The discovery of these effects was one of far-reaching significance; it laid the foundation stone for two new sciences, namely, for the science of *X-ray spectroscopy*, on the one hand, and for the *analysis of crystal structure* as now pursued, on the other hand.*

The experimental arrangement of Friedrich and Knipping is represented in Figure 195*a*. The X-rays originate at the target *T*, which is bombarded

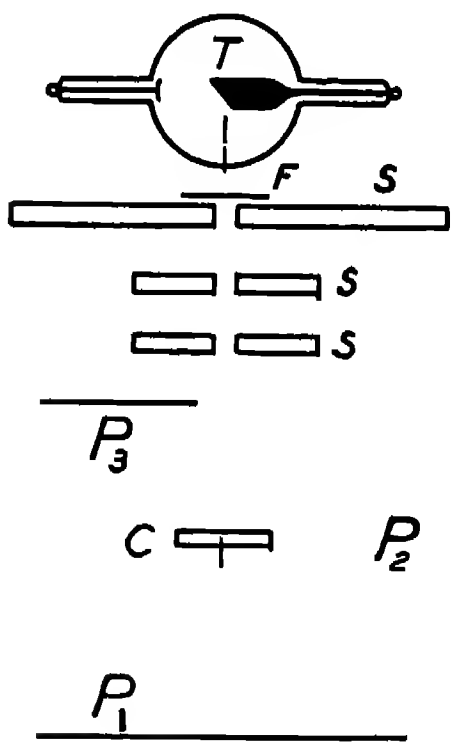


FIG. 195*a*

FIG. 195*b*.—Laue spots

by cathode particles coming from the cathode shown at the left (not lettered). The X-rays pass through a filter, *F*, of aluminum, and through circular apertures about 1 mm in diameter in a succession of lead screens,

* The original article describing the discovery is: Friedrich, Knipping, and Laue, *Sitzungsber. d. Muench. Akad.*, p. 303, 1912. This was reprinted with only minor alterations in *Ann. d. Phys.*, **41**, 971, 1913.

SS, and fall upon the crystal, C, which was about 0.5 mm thick in the experiments in question. Photographic plates were placed at P_1, P_2, and P_3, some 3.5 cm from the crystal. The plate at P_1 yielded a pattern of the type shown in Figure 195*b*. The large central spot is due to the direct beam of X-rays which continues straight through the crystal. The smaller surrounding spots are the images formed where the directions representing the principal maxima of diffraction penetrate the photographic plate. These spots are now known as "Laue spots." Other Laue spots appeared upon the plates at P_2 and P_3.

The Laue theory of the diffraction of X-rays by crystals, while based upon the simple principle of the interference of the wavelets scattered by the atoms of the crystal, is nevertheless rather difficult to grasp because we must visualize the interfering wavelets in three-dimensional space. But soon after Laue published his theory, W. L. Bragg found a simpler viewpoint from which to regard the diffraction of X-rays by crystals. When we are ready to undertake the consideration of this diffraction, we shall begin with Bragg's method of viewing it, but we must first digress to consider some of the elementary features of crystal structure. Some of these features were matters of mere conjecture before Laue's discovery and Bragg's interpretation of this discovery made it possible to analyze crystals by means of the X-ray diffraction effects which they produce, but these features may now safely be considered as definitely known.

4. Concerning Crystal Structure.—In every type of space lattice there is a so-called "unit cell" which is a geometric volume, a unit parallelepiped, of which the lattice may be supposed to be built up as of building blocks. In general, the edges of the unit parallelepiped may have different lengths and be inclined to each other. Each corner of the parallelepiped is considered to be a "lattice point." In simple crystals it is usually convenient to consider the lattice points as coinciding with the atoms. But in complicated structures several atoms may be considered to "center" about a lattice point which does not necessarily coincide with any one atom. The lattice itself is then a geometrical abstraction serving for reference, having form and dimensions determined by the structure of the actual crystal. When it is possible to choose a parallelepiped whose edges make right angles with each other and all have the same length, then the unit cell is a cube, and the crystal belongs to the cubic system. In this case there is a single "lattice constant," namely, the length of the edge of the unit cube. This constant will be denoted by the symbol a. Crystals of sodium chloride, rocksalt, for example, belong to the cubic system, and it is now known that the atoms are arranged as shown in Figure 194. There, the solid circles may be taken to represent the Na atoms and the empty ones the Cl. The heavy connecting lines outline the unit cube of Na atoms; the length a is indicated

in the figure. Denoting the distance from one atom to its nearest neighbor by e, that is, the distance from Na to Cl or vice versa, *like* atoms are separated by a distance $a=2e=5.628$ Å along the edges of the cube. Several further features are to be noted: The unit cube of Cl atoms is displaced, with reference to the cube of Na atoms, by a distance e, upward in our figure, and penetrates the cube of Na atoms. The unit cube of each kind of atoms repeats itself identically at the interval a measured along any one of the cube edges. The cube formed by Na atoms has atoms not only at the corners of the cube, but also an Na atom at the center of each face. The cube formed by Cl atoms has likewise a Cl atom at each corner and at the center of each face. The lattice is said to be of the "interpenetrating face-centered" type.

(The reader will find a model of a cubic lattice, if he can procure or improvise one, very helpful in understanding the present as well as following discussions.)

In a crystal, the molecule, as such, loses its identity. For there is no reason for saying that a given Na atom forms a molecule in conjunction with the neighboring Cl atom to its right, rather than with the Cl atom to its left, or the one above, or the one below.

While the molecule as such loses its identity, the content of an integral number of molecules is always associated with each unit cell; hence in this sense we may say that each unit cell comprises such-and-such a number of molecules. Considering the cube outlined by the heavy connecting lines, of length of side a, in Figure 194, let us associate with this unit cube the eight atoms—four of Na and four of Cl—which lie at the corners of the smaller cube of length of side e in the front lower left-hand corner of the larger cube indicated by the heavy connecting lines. Making this association, all other atoms lying at a corner, upon an edge or a face of the larger cube, become, by the same scheme, associated with one or another of the adjacent unit cells. For example, the eight atoms at the corners of the small cube at the front upper left corner of our figure become associated with the next unit cell above the one given by the heavy connecting lines. In the same way all other atoms become associated with other unit cells. There are thus four molecules, and four only per unit cell.

There are many other cubic crystals of the face-centered type. Among them are, for example, copper, silver, and gold. These latter crystals have, however, only a single kind of atom, and the atoms are arranged in a single lattice, not in two interpenetrating ones. But diamond, made up also of only one kind of atom, is cubic and yet consists of two interpenetrating face-centered lattices displaced along the body diagonal of the cube by one-quarter of its length. Further, the crystal zinc blende, zinc sulphide, is made up of two kinds of atoms but otherwise resembles diamond. Diamond and

zinc blende were among the crystals used in the original investigation of Friedrich, Knipping, and Laue. Some elements—for example, lithium, sodium, and potassium—crystallize in the "body-centered" cubic form. Attention is called to these various facts in order to point out circumstances which must be considered in the study of cubic crystals—and the structure of crystals belonging to systems other than the cubic is still more complicated.

The arrangement of particles shown in Figure 193 is known as a "simple cubic" lattice. This is an *ideal* system to which no actual crystal corresponds. But *if*, in a crystal of the rocksalt type, two different kinds of atoms scatter X-rays to about the same degree, the structure *may be regarded* as simple cubic by way of close approximation. When so regarded the unit cube having, by definition, a length of side a, reduces to one of side e, where e is by definition the distance from one atom to its nearest neighbor, that is, in this case $a=e$.

In studying space-lattice diffraction it is of importance to understand first of all the effects produced by the simplest of all lattices, namely, the simple cubic, and this type will be taken as a basis in much of our subsequent discussion. It is best exemplified by crystals of sylvite, K Cl. The structure of sylvite is the same as that of rocksalt, NaCl, but in sylvite the K and Cl atoms scatter almost exactly equally for a reason which we shall explain. The scattering due to any one atom is attributable largely to the electrons outside of the nucleus of the atom. The constituents of the nucleus are too firmly bound to be effective, and the nucleus as a whole does not contribute materially, since, on account of its large mass, it reacts very little to the incident wave train and hence does not scatter the waves appreciably. Now the number of electrons outside of the nucleus of a potassium atom in its elemental state is 19, and the corresponding number for chlorine is 17. But when the elements potassium and chlorine are in chemical combination, as in the sylvite crystal, they form a so-called "polar compound." That is, the single valence electron of potassium associates itself with the seven valence electrons of chlorine. The lattice particles are thus, respectively, positive and negative ions; the lattice is said to be of the "ionic" type. The effective number of electrons outside of the nucleus for each of the elements K and Cl when in chemical combination is thus 18. And the result is that the crystal sylvite behaves very nearly as would a simple cubic lattice.

Our concept of the simple cubic lattice requires that the particles must all be alike in every respect and, in particular, all have the same size and shape. The shape may, however, be an arbitrary one without symmetry of any kind. But the particles must all be similarly oriented. To emphasize

these facts, the particles are sometimes depicted in the form of heavy inverted commas. Further, we do not need to concern ourselves with the details of the size and shape of the lattice particles at all while we are seeking to ascertain the *directions* of the maxima of diffraction, just as with a grating we do not need to concern ourselves with the details of the width or form of the grooves in ascertaining the directions of the maxima formed by it. Accordingly, in determining the directions of the maxima, we may for simplicity regard the atoms of a crystal as concentrated at geometrical points, although actually, with their surrounding electrons, they extend over a considerable fraction of the lattice space. However, just as, with the grating, the groove form becomes of importance in determining the relative *intensities* of the maxima which are formed, so, analogously, the "atom-form factor" or "structure factor" becomes important whenever questions of the intensity distribution of X-ray diffraction or scattering arise. Something more concerning structure factor is said in section 6.

In elementary theory we may also make the simplification of regarding the atoms as *at rest,* whereas they are in fact in a state of agitation due to temperature.

Any plane of a crystal which is occupied by lattice points is known as a "crystal plane" or "net plane," the latter because of the netlike arrangement of the points in such a plane. A crystal face is always parallel to a set of net planes.

Referring again to Figure 193, consider the plane parallel to the Z-axis which makes an intercept a upon the X-axis and $2a$ upon the Y-axis—the intercept of this plane with the Z-axis is of course infinity. The so-called "intercept ratios" of this plane are $1:2:\infty$. All net planes parallel to this plane would have the same intercept ratios and these planes may each, or all, as a set, be specified by the intercept ratios $1:2:\infty$. It is only the orientation of the planes with which we are concerned—not their absolute positions. However, the orientation of planes is not usually specified by their intercept ratios, but commonly instead, by their so-called "Miller indices," which are a set of integers in the ratio of the reciprocals of the intercepts, namely, in the ratio $1:\frac{1}{2}:\frac{1}{\infty}$ in our example. The numbers of this ratio are multiplied by the smallest integer, 2 in our example, which will serve to convert each of these numbers into an integer, and the integers so obtained are the Miller indices or simply "indices," which are thus a set of integers which have no common divisor, or are prime to each other, as a set. The indices are simply written in succession. They are commonly, but not always, placed in parentheses, thus (210); comas are inserted between the numbers when for some reason necessary for clearness. By way of further example: The set of planes perpendicular to the X-axis has intercept ratios

$1:\infty:\infty$, and indices (100); the planes parallel to the Z-axis and intersecting the X- and Y-axes at 45° have intercept ratios $1:1:\infty$, and indices (110); the plane intersecting each of the co-ordinate axes at a distance a from the origin, and all planes parallel to this one, have intercept ratios $1:1:1$ and indices (111).

A crystal face having indices (100) is a so-called "cube face," and crystal faces having indices (010) or (001) are also cube faces. From a physical point of view there is no distinction between these faces, so far as the crystal itself is concerned. It is only when, for example, a beam of X-rays is incident in the direction of a given cube edge—that is, in the direction of a given co-ordinate axis—that a distinction must be made between different cube faces, for example, between (100) and (010). In the same connection it may also be necessary to distinguish between the planes having intercept ratios $1:2:\infty$, or indices (210), and the planes having ratios $1:-2:\infty$, or indices $(2, -1, 0)$.

Each index number is always given the algebraic sign of the corresponding intercept. Negative signs are usually written above the index number, thus $(2\bar{1}0)$. In passing from one plane to a parallel plane on the opposite side of the origin, each index number reverses in sign. Apart from this reversal in sign the planes of a parallel set all have the same indices.

Intercept ratios and Miller indices are further discussed in Appendix G, but what has already been said is sufficient for our immediate purposes.

5. The Bragg Theory.—We are now ready to consider the view of space-lattice diffraction which was propounded by W. L. Bragg* soon after Laue announced his discovery.

Bragg considered the diffraction which occurs, from the viewpoint of *reflection*, in the following manner: Suppose a beam of X-rays is incident upon a lattice in the direction of the arrow i at the top of Figure 196. The incident beam may be regarded as parallel; it cannot be rendered parallel by a lens or mirror as an optical beam might be, but we may regard it as parallel nevertheless so long as we are considering the effects which arise from only a small portion of the lattice at one time. This is true because the distance between the lattice particles is so extremely small. In the scale of magnification to which our figure is drawn the rays should be considered to originate from a source point a thousand miles distant. And, incidentally, in the same scale, a different surface element of the crystal might be forty or fifty miles distant. At this other point the waves could again be regarded as plane but the angle of incidence would be different.

In view of the circumstances which obtain, the straight broken line of our figure which is perpendicular to the arrow i may be taken to represent an

* *Proc. Camb. Phil. Soc.*, **17**, 43, 1912.

incident wave front which may be regarded as being plane. As this wave front strikes the atoms of the crystal in succession, an elementary wavelet originates from each atom in turn. Let us consider the wavelets which originate from the first layer of atoms *ff*. Carrying out the familiar Huygens construction, it is evident that these wavelets combine into a plane wave, W_1, which travels in the direction of the arrow *d*, this direction making an angle with the plane of the layer *ff* equal to that made by the direction *i* with this plane. The wavelets thus combine into a wave front which may be spoken of as "reflected" by the layer of atoms. Let us now consider the wave front which is built up by the wavelets scattered by the second layer of atoms *ss*. These wavelets combine into a wave front W_2 which also travels in the direction *d*, but this wave front is retarded with reference to the wave front W_1 by a certain amount δ. Let us evaluate the retardation δ: We may regard both the incident wave and the wavelets from the second or lower layers as traveling through the crystal with the same velocity as in air or vacuum. If the velocity of X-ray waves within matter differed appreciably from the velocity in vacuum, X-rays would be refracted appreciably upon entering or leaving matter. And while it is true that they are refracted somewhat, the amount of refraction is so extremely small that it long eluded detection completely, and was eventually discovered only by the most refined and careful measurements. These measurements, and other lines of evidence which we need not now discuss, led to the conclusion that the *wave* velocity of X-rays is *greater* in matter than in vacuum, but only by about 1 part in 10,000 even in rather dense solids. (The *group* velocity cannot exceed the velocity in vacuum.) Ignoring then as we well may, for the present, the effect of this minute change of velocity which occurs, it is evident that the retardation δ has the value:

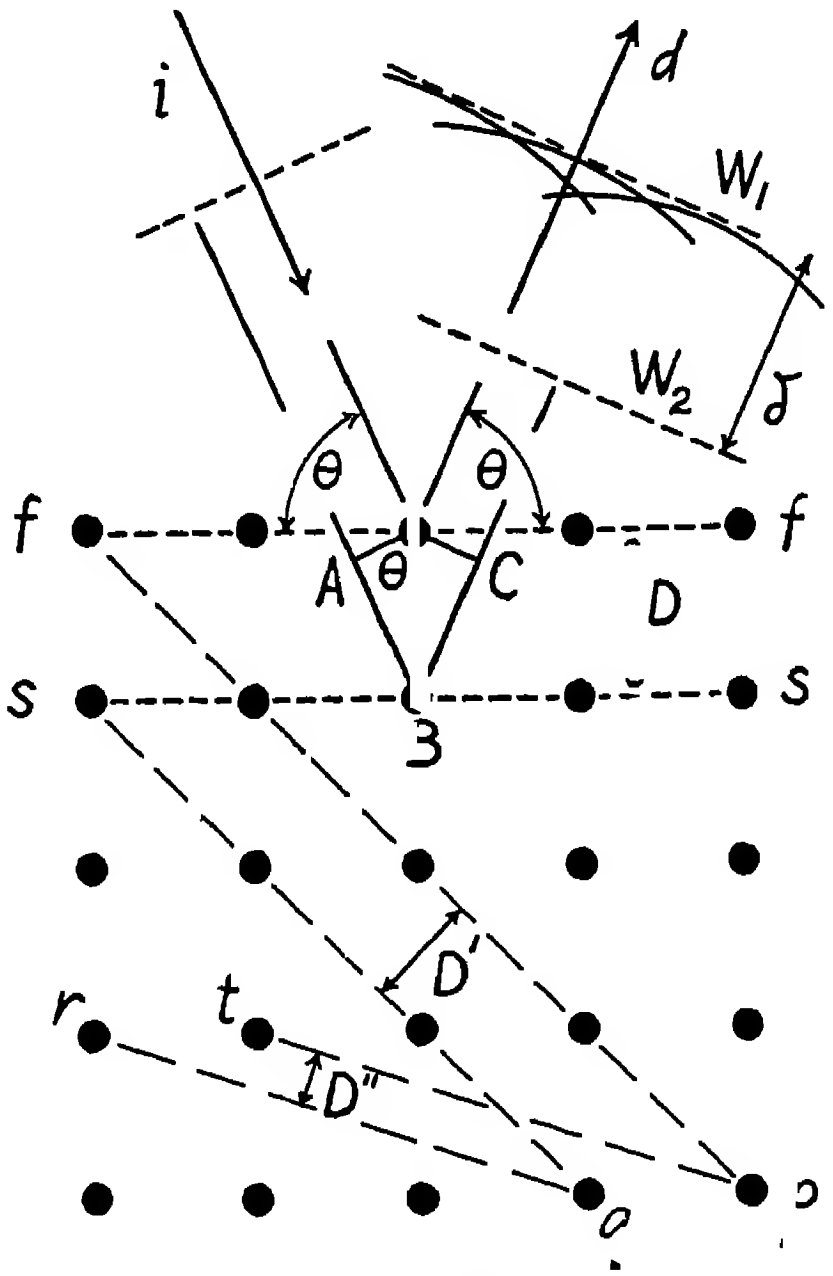


FIG. 196

$$\delta = AB + BC = 2D \sin \theta \,. \tag{2}$$

Here D is the distance indicated in the figure and θ is the angle between the direction of incidence and the planes of the layers *ff*, *ss*, etc., and this angle is equal to the angle between the direction of diffraction and the planes of the layers. The angle θ is known as the "glancing angle" made by the beams with the layers. The "glancing angle" must always be distinguished from the "angle of incidence" which is the angle between the incident beam and the *normal* to a surface. The distance D may be defined in general as the space period in which similar layers of lattice particles recur, measured perpendicularly to any set of layers which may be in question. The fact that D in the present instance equals our previous e, the distance from one lattice particle to its nearest neighbor, or a, the lattice constant, is an incidental circumstance, as will appear more fully later.

The Huygens wavelets which originate from the third layer of atoms, the fourth layer, etc., also combine into wave fronts traveling in the direction of the arrow d. Moreover, the wave front originating from each successive layer is retarded relatively to the wave front from the preceding layer by always the same amount δ, given by equation (2). Now suppose the incident radiation contains trains of waves of all wave-lengths. The disturbances from the various layers of atoms will add up—that is, interfere constructively—only for those wave-lengths for which the retardation δ equals an integral number of wave-lengths, $m\lambda$. For all other wave-lengths there is practically complete destructive interference, just as with a diffraction grating there is practically complete destructive interference for all wave-lengths except those for which the grating law is satisfied in the given direction. Remembering, then, that δ must equal $m\lambda$, we pass from equation (2) to what is known as Bragg's law for X-rays "reflected," or really diffracted by a crystal, namely:

$$m\lambda = 2D \sin \theta . \tag{3}$$

Here m is the spectral order. In any given order the wave-length which is diffracted increases as the glancing angle θ increases. For any given wave-length the first order occurs at the smallest glancing angle θ_1, the second order occurs at a greater angle θ_2, the third order at a still greater angle θ_3, etc. The low orders are the ones of practical importance because the intensity of diffraction falls off rapidly with increasing order for a reason which we shall learn in the next section.

Bragg's law, taken literally, states that a given wave-length will be reflected in a given spectral order only when the glancing angle takes a *unique* value θ. Our instincts regarding physical phenomena lead us, however, to surmise that actually there will be a small "angle of tolerance" over which diffraction will occur, this angle being finite, however small it may

be. And such a small finite range of angle does indeed exist—speaking generally, it has an order of magnitude of several seconds of arc. In sections 21 and 24 we shall learn more regarding the magnitude of this small angle.

It is to be noted that in order to have reflection from a crystal occur it is not *necessary* that the lattice particles should be regularly arrayed *within* each of the reflecting planes. Referring again to Figure 196, suppose any one of the lattice particles were displaced by an arbitrary amount within its own plane *ff*, *ss*, etc. The wavelet originating from it would still be tangent to the same wave front as before. Therefore, if the lattice particles occupied random positions within their own planes instead of being regularly arrayed within them, the wavelets would still build up the same wave fronts. The formation of the reflected beam is thus conditioned upon the existence of equally spaced layers of particles, but not upon the regularity of arrangement of the particles within these layers. The regularity within each plane is incidental in so far as reflection in the given set of planes is concerned.

The regularity which actually obtains within each set of layers means, however, that it will be possible to find other, differently oriented sets of planes which are also occupied by lattice particles—for example, the set *fp*, *sq*, etc. (Fig. 196)—and these other sets may equally well be regarded as constituting "reflecting" planes.

Since a fixed number of lattice particles, or atoms, per unit volume of a given crystal must always be accounted for, those sets of planes which have many atoms per unit element of area, or are "densely populated," have a relatively large separation between planes. Such sets are *ff*, *ss*, etc., or *fp*, *sq*, etc., the corresponding separations being respectively D and D'. And those sets of net planes which are sparsely populated, such as *tp*, *rq*, etc., have a correspondingly small separation, D''.

Those sets of net planes which are densely populated commonly occur as crystal faces. For these planes the Miller indices are "simple." That this is true is more or less self-evident, but the question which here arises is treated mathematically in Appendix G. The planes which are very sparsely populated have, on the other hand, Miller indices which are not simple, and these planes do not occur as faces. Furthermore, the densely populated planes serve well as "reflecting planes," whereas the "reflection" from the sparsely populated planes is of low intensity. But the reason for this low intensity of reflection does not lie immediately in the sparsity of population, for there is always a constant number of scattering atoms for a constant volume of penetration of the beam. The reason for the low intensity is that the separation of the planes is so small that the extent in space of the atoms, with their surrounding electrons which cause the scattering, is relatively so great that the distribution of the scattering electrons, considered in the direction perpendicular to the planes, somewhat approaches uniformity.

Temperature agitation of the atoms depresses the intensity of reflection from the sparsely populated planes also because of the small separation which these planes have.

To sum up, the densely populated planes are the ones which are important both crystallographically and spectroscopically, whereas the sparsely populated planes are unimportant. The sparsely populated planes are unimportant *physically*.

Those sets of planes which are parallel to two of the principal edges of the lattice are densely populated, but a set of planes may be inclined to all three of the principal edges of the lattice and yet be densely populated, as, for example, the (111) set, and hence be important physically.

It is to be noted that the following conditions obtain in the simple cubic lattice, with which we are especially concerned:

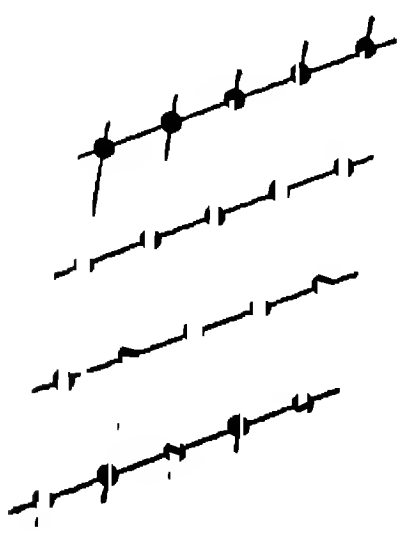


FIG. 197

All lattice particles are exactly alike, and all have neighbors in the same relative positions.

The most general type of arrangement which is possible in any one net plane is that shown in Figure 197. No other type satisfies the previously mentioned conditions.

When a set of planes is inclined to one or more of the cube edges of the lattice, the atoms in adjacent planes do not "lie behind each other"—but this is immaterial from the point of view of reflection, as we have learned.

By passing from a given plane to the next but one, or to the third, or the fourth, etc., it is always possible to find a plane in which the atoms lie exactly behind those of any plane originally considered.

6. Some General Aspects of Space-Lattice Diffraction.—A fundamental distinction exists between diffraction by a space lattice and diffraction by a grating. With a grating, for a given direction of incidence, there is a continuous range of directions in which a continuous range of wave-lengths is diffracted in a given spectral order. There is, in other words, a "spectrum" formed. With a space lattice, for a given direction of incidence, there is only a single direction of diffraction (for a given set of reflecting planes), and only a single wave-length is diffracted or "reflected" in a given order. There is thus no "spectrum" formed, when there is a single direction of incidence. But if a *continuous range of directions of incidence* is provided by having a divergent incident beam, a portion of a spectrum will be formed at any one crystal setting; details will be explained later, and it will also be explained how a complete spectrum may be obtained by rotating the crystal either slowly and continuously or by small steps.

Let us now see why the beams of successive order diffracted by a crystal fall off rapidly in intensity as the order m increases. The reader will recall that in diffraction by a grating the amplitude of the disturbance which prevails at any principal maximum is proportional to the amplitude which each individual groove of the grating sends out in the direction in which the given maximum is formed. Similarly, in diffraction by a crystal, the amplitude of any principal maximum is proportional to the amplitude of the disturbance which the individual atom scatters or diffracts in the given direction. Now the amplitude contributed by the individual atom decreases rapidly with increasing order—for a reason which will appear.

The higher the spectral order, the greater the glancing angle θ which the incident and diffracted beams make with the reflecting planes, and, also, the greater the angle through which the diffracted beam is deviated with relation to the incident beam. Denoting this angle of deviation by ψ, we have specifically $\psi = 2\theta$. Now the amplitude of the disturbance scattered by each atom in any direction decreases as the angle of deviation ψ increases, because as ψ increases the interference between disturbances scattered by different electrons of the same atom tends progressively to become more destructive, since the path differences by way of different parts of the same atom increase progressively. Referring to Figure 198, let the arrow i represent the direction of an incident beam of X-rays of wave-length λ. Let the broken line perpendicular to the arrow represent a wave front, and let the circle bf represent the space occupied by an atom, with its surrounding electrons which cause the scattering. Within the atom each element of volume, small compared to the wave-length, may be regarded as scattering without great variation with direction, much as in the Rayleigh scattering of light by a particle small compared to the wave-length when the incident beam is unpolarized. But the scattering due to the entire atom will vary markedly with direction. In the direction d' which is but little deviated from

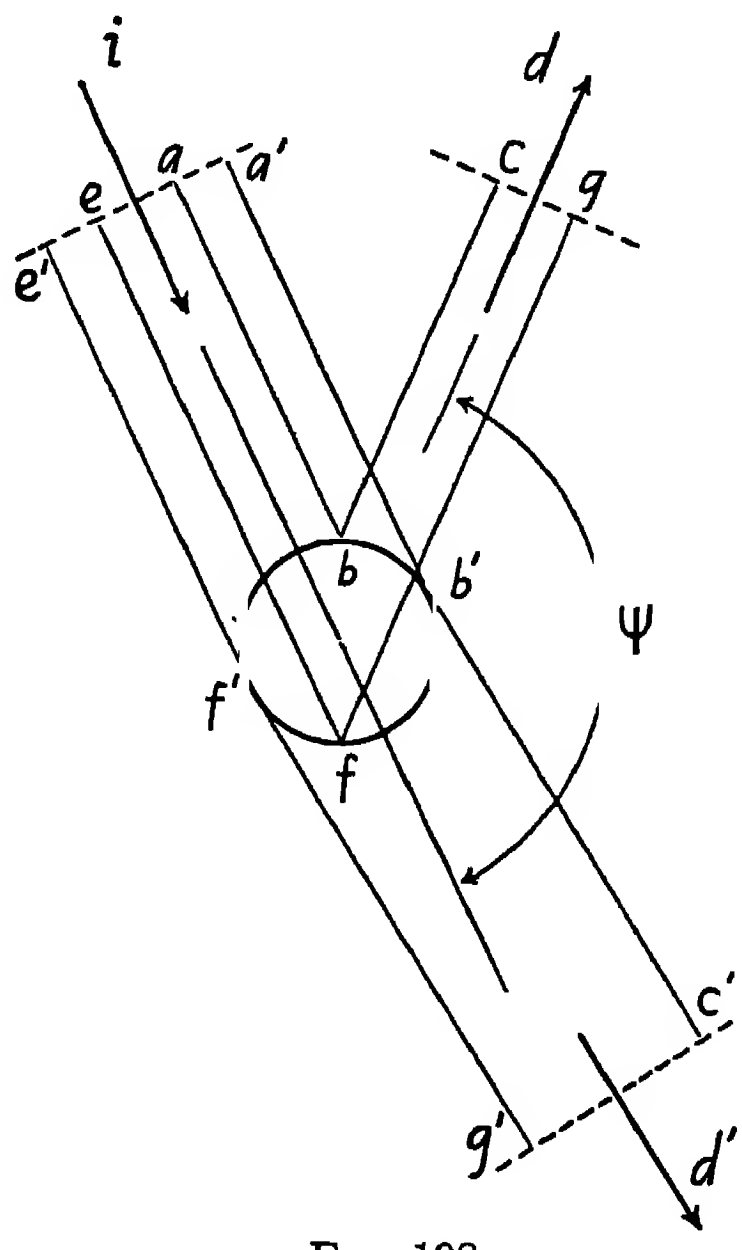


FIG. 198

the direction of the incident beam, the difference between the paths $a'b'c'$ and $e'f'g'$, by way of extreme parts of the atom, will be small. Supposing that the deviation of d' with reference to i is sufficiently small, the paths in question will differ by only a small fraction of a wave-length and the disturbances scattered by the various electrons of the atom will reinforce each other almost perfectly. But in the direction d which is deviated by a large angle ψ from the direction of incidence, the extreme path differences abc and efg may differ by a large fraction of a wave-length or by even considerably more than a wave-length, depending upon how large the atom is in relation to λ. Hence the disturbances scattered by the various electrons distributed throughout the space occupied by the atom may almost annul each other mutually in the direction d. Thus the resultant disturbance scattered in any direction by an atom decreases with increasing angle of deviation ψ. And the reason for this decrease is because the atom is not concentrated at a geometrical point but has finite extent in space.

The so-called atom structure factor is defined as being equal in any given direction to the ratio of the amplitude of the disturbance scattered in the given direction by all of the electrons of an atom to the amplitude of the disturbance which would be scattered in the same direction by a single electron on the basis of classical theory. Now in the direction of the incident beam continued—that is, in the direction of zero deviation—the disturbances scattered by the various electrons of an atom will reinforce each other perfectly. Hence the amplitude of the disturbance scattered in this direction will be Z times the amplitude scattered by a single electron, where Z is the atomic number of the atom. With increase in the angle of deviation, starting from zero, destructive interference progressively develops. Accordingly, the structure factor has a maximum value of Z in the direction of zero deviation and decreases with increasing angle of deviation at a rate which depends on the wave-length of the incident beam.*

We have now learned that one reason why the intensity falls off rapidly in passing to higher orders is because of the finite extent of atoms in space. Orders higher than the sixth or seventh are ordinarily not found at all. (There may, however, be special circumstances which cause certain high orders to appear.†)

* For further discussion of the subject of structure factor see, e.g., E. O. Wollan, *Rev. of Modern Phys.*, **4**, 205, 1932.

† Some organic crystals of long-chain compounds have a repeating structure within the molecule. The distance D between identical net planes equals the length of the molecule, but numerous recurrences of the same atom group within the identity interval D cause certain higher orders of diffraction to be reinforced. With the crystals in question, orders higher than even the sixtieth may be observed. See, e.g., A. Mueller, *Proc. Roy. Soc.*, **120**, 437, 1928.

Another cause for the suppression of the higher orders in relation to the lower ones is temperature agitation of the atoms within a crystal. When the atoms are in a state of agitation the effective thickness of the atom layers is greater than if the atoms were at rest. This increase in effective thickness is less damaging at small angles of deviation—that is, for low orders—than at large angles of deviation, or for higher orders, and the result is much the same as though the atoms were larger than they actually are.

Let us now inquire what the range of wave-lengths is which may be diffracted by crystals. From Bragg's law, $m\lambda = 2D \sin \theta$, we learn that when the glancing angle is small, the measured wave-length is small, and that the wave-length increases with increasing glancing angle, up to a maximum value at normal incidence, $\theta = 90°$, where, with $m = 1$, we have $\lambda = 2D$. The maximum value which D can have is a, the lattice constant of the crystal, hence, in analogy with diffraction by a ruled grating, the maximum wave-length which can be diffracted by a crystal is $\lambda = 2a$, or a wave-length equal to twice the lattice constant of the crystal. From a practical point of view, the range of wave-lengths which may be measured by diffraction from crystals extends from a lower limit of .05 or .1 Å to an upper limit of some 15 or 20 Å. The lower limit is set by a combination of factors. Let us take an example. Let $m = 1$, $\lambda = .1$ Å, and $D = 2$ Å. Applying Bragg's law, we have $1 \times .1 = 2 \times 2 \times \sin \theta$ or $\sin \theta = .025$. That is, the glancing angle is required to have the small value of about 1°.5. At this same glancing angle the second order, $m = 2$, yields $\lambda = .05$ Å, but in passing to the second or higher orders there is a sacrifice of intensity. Further, the above-given separation of reflecting planes $D = 2$ Å is about as low a value as can be attained in practice. Of all crystals, diamond has the smallest lattice constant, namely, $a = 3.55$ Å. By reflecting from planes of the diamond crystal which are equally inclined to the three cube edges, the value of D becomes $3.55/\sqrt{3} = 2.05$ Å, and reflection from sets of planes having smaller values of D would be accompanied, again, by a sacrifice of intensity. Now as to the upper limit of the range of wave-lengths which may be measured by diffracting from crystals: Gypsum ($CaSO_4 + 2H_2O$) has a lattice constant $a = 7.62$ Å; accordingly, one would measure the wave-length $\lambda = 2a = 15.24$ Å in the first order at normal incidence. Gypsum has the largest lattice constant of all stable crystals except mica. With mica, wave-lengths up to nearly 20 Å may be measured. There exist certain unstable crystals formed by organic compounds which have much larger lattice constants than gypsum or mica, and, with these, wave-lengths up to 60 Å have been measured upon occasion, but the use of such crystals is not practical. The ruled grating is more suitable for the measurement of the wave-lengths of very soft X-rays.

Another question which arises in connection with crystal diffraction is

whether by considering the beams which are reflected, according to the Bragg view, in the various sets of net planes, we are taking account of all of the principal maxima of diffraction which may be formed, or whether other principal maxima may be formed which cannot be accounted for as being "reflected." Actually, the Bragg view does account for all of the principal maxima which are formed, and we shall presently explain in a general way why it does. A more formal proof that it does so is submitted in Appendix G, but this proof assumes that the reader has first studied the Laue theory of space-lattice diffraction, which, in its very nature, determines all of the directions in which principal maxima are formed.

Referring to Figure 199, let i be the direction of an incident beam of wave-length λ, and let d be the direction of any principal maximum of diffraction for this wave-length. When the Huygens wavelets which originate from the particles of the lattice all reinforce each other, as they must when we are dealing with a principal maximum, they build up wave fronts W, W', etc., which are separated by a distance λ. Let A be a given lattice point. We may, for simplicity, think of A as being a single atom, though more generally it may represent a group of atoms which center about this position. Let A, at a certain instant of time, give rise to a Huygens wavelet which aids in building up a wave front which at a given later instant will be located at W. Erect a plane rr through A which makes equal glancing angles, θ, with the incident and diffracted beams. Obviously all other lattice particles which at the same original instant gave rise to wavelets which aid in building up the wave front W must lie somewhere in the plane rr. Further, if other lattice particles contribute Huygens wavelets which build up the wave front W', these particles must lie in a parallel plane $r'r'$ which is separated from rr by a distance D given by Bragg's law when $m=1$. Thus the set of planes rr, $r'r'$, etc., may be regarded as "reflecting" planes for the diffracted beam, which is *any* diffracted beam.

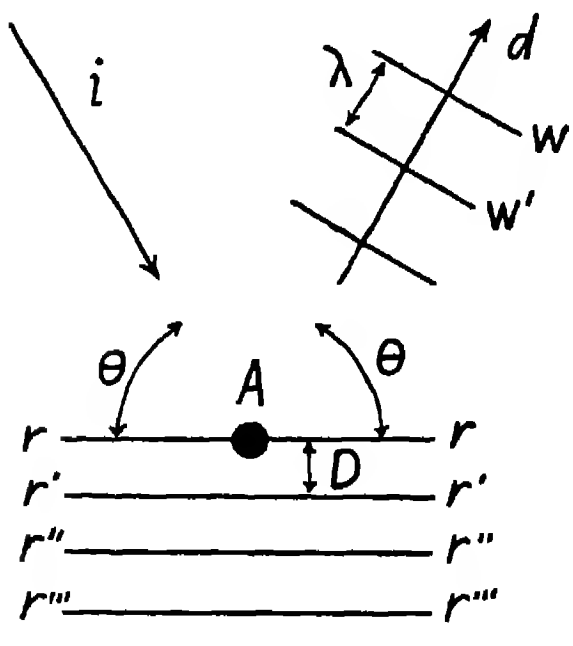


FIG. 199

All of the planes of the set rr, $r'r'$, etc., do not need to be occupied by lattice points. For example, the planes rr, $r''r''$, etc., might be occupied and the planes $r'r'$, $r'''r'''$, etc., be unoccupied. In this case the separation of the reflecting planes becomes twice the D of our present figure, and we are dealing with second order reflection in a set of planes of separation $2D$.

7. Spectrographs for Soft X-rays.—First, a few words concerning X-ray tubes: Formerly, the cathode of the tube was simply a small round sheet

of metal either flat or concave. Gas at a pressure of one or more thousandths of a millimeter remained in the tube, and when the potential was applied, positive ions were formed in this residual gas. These positive ions bombarded the cathode and caused it to give off electrons which then traveled to the anode or target of the tube, and, in impinging upon this, generated the X-rays. The presence of the residual gas was thus a necessity for the functioning of the tube. In the course of use, the gas became partially absorbed in some manner by the walls of the tube, and the tube was then said to have become "hard." In modern tubes, on the other hand, the cathode consists of a tungsten wire, usually wound in the form of a spiral, which is heated to incandescence by an electric current supplied by a storage battery or other source of potential which is independent of the main source of potential which actuates the tube. The incandescent tungsten wire or "hot cathode" serves as a copious source of electrons, and it is hence no longer necessary to have residual gas in the tube. In fact, the tube should, for several reasons, remain highly evacuated. A short cylinder of molybdenum surrounds the hot cathode. This cylinder serves the purpose of concentrating or "focusing" the cathode rays upon the anode or target of the tube. The hot cathode tube represents an important development in X-ray technique over the former "cold cathode" tube because it is much more nearly constant in performance and admits of much better control of the current as related to the voltage applied. To dissipate the heat which is generated, the anode, and other parts of the tube also, are frequently water cooled. The hot cathode tube was invented in the year 1913 by W. D. Coolidge.*

In investigating the spectra due to soft X-rays, the X-ray tube is usually built directly on to the spectrograph. A thin membrane is commonly introduced to separate the two chambers, and the tube and spectrograph are each evacuated. This procedure avoids the absorption of the radiation which would otherwise occur in the wall of the tube and the absorption due to the air. With X-rays of wave-length greater than two or three Ångstroms, such absorption must be avoided.

Referring to Figure 200*a*, let T be the target of the X-ray tube and S a slit having jaws usually made of brass, which is sufficiently dense to be opaque toward soft X-rays. Let C be a crystal and PP' a photographic plate. Denote the glancing angle which the ray SC makes with the crystal surface by θ, and suppose the direction CP makes the same glancing angle with this surface. There will thus be diffracted, or "reflected," to P a wave-length λ, in the order m, given by Bragg's law $m\lambda = 2D \sin \theta$. Now the depth of penetration of soft X-rays into the crystal will not exceed a few hundredths of a millimeter. Consequently, while the depth of penetration

* *Phys. Rev.*, **2**, 409, 1913.

comprises a multitude of layers of atoms, it is yet so small that we may, when the principles of design of a spectrograph are under consideration, regard the "reflection" as occurring at the geometrical surface of the crystal.*

Since the target, T, is extended in area, the beam issuing from the slit will be divergent, and hence the glancing angle will vary from point to point

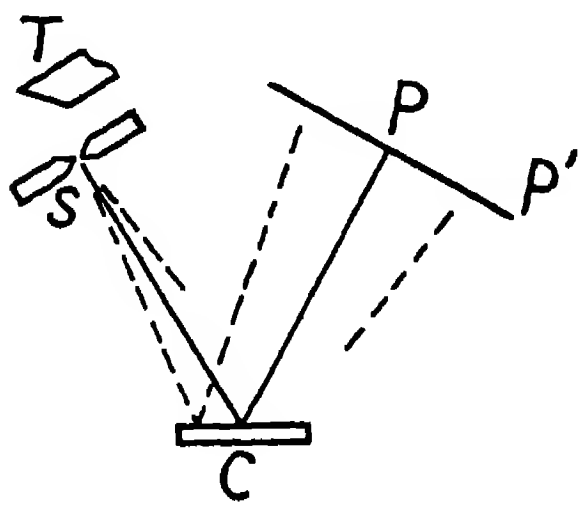


FIG. 200*a*

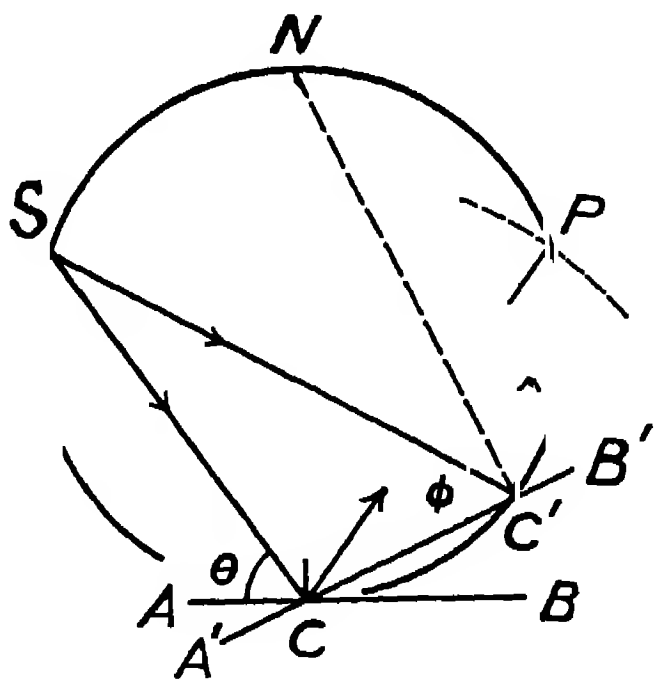


FIG. 200*b*

of the crystal surface over a small range, and consequently the wave-lengths which are diffracted will vary over a small range. It will thus be possible, owing to the fact that the incident beam is divergent, to photograph a small portion of the spectrum at any given angular setting of the crystal, and the whole spectrum may be explored by photographing in a succession of settings of the crystal and photographic plate. This is, in outline, the method of the diverging beam; it was originated by H. J. G. Moseley† in his renowned investigations upon X-ray spectra which first definitely disclosed the importance of assigning to each chemical element an "atomic number."

Referring still to Figure 200*a*, let us now for a moment consider only the ray SC of the incident beam and suppose that the crystal is rotated slowly about an axis C which lies in the plane of the reflecting surface and is perpendicular to the plane of our figure. The glancing angle will vary, and hence the wave-length which is diffracted will vary. Consequently, even if only a narrow pencil of rays were incident, a spectrum would be formed upon the photographic plate. This is the so-called "rotation method," and it may be combined with the method of the diverging beam under circumstances which we shall explain shortly.

When the diverging-beam method is used without rotating the crystal, falsifications are often introduced into the spectrum. These may take the

* An exception occurs in the case of crystals of very low density, primarily organic crystals, e.g., sugar. With these the depth of penetration is considerably greater.

† *Loc. cit.*

form either of false spectrum lines or of falsification of the intensity distribution in the continuous spectrum. These falsifications arise from small areas of the target which yield especially intense radiation and from differences in the intensity of diffraction from various portions of the crystal surface. The method of using a diverging beam and at the same time rotating the crystal, which we shall now describe, smooths out the effects of these inequalities.

If the distance PC is made equal to SC, and the crystal is rotated about an axis C which lies in the reflecting surface and is perpendicular to the drawing, the point of incidence of that ray which is at any given time "reflected" to P will move across the crystal, but the glancing angle will remain constant—as will be proved—and hence the wave-length which is diffracted to P will remain always the same while the crystal is rotated. Referring now to Figure 200*b*, let S denote the position of the slit, C the axis of rotation of the crystal, and P any given point lying on an arc described about C with radius $CP = SC$. Pass a circle through the points S, C, and P. Let AB be the position of the crystal such that the ray incident at C is reflected to P, and denote the glancing angle which then obtains by θ. The normal to the crystal face erected at C intersects the circle SCP at N. Now suppose that the crystal is rotated to any other position $A'B'$, in which the crystal surface intersects the circle at C'. Denote the glancing angle at C' by ϕ. Erect a normal at C'; this normal cuts the circle SCP at N and $\angle SCN = \angle SC'N$. Hence the glancing angles θ and ϕ are equal, and hence the wave-length "reflected" to P remains always the same while the crystal is rotated.

Thus, the rotation method and the method of the diverging beam may be used in combination, provided the photographic plate and the slit are both placed at the same distance from the axis of rotation of the crystal. This is the so-called "focusing condition." To fulfil this condition accurately for all points of the photographic plate, the plate should be bent to an arc about the axis of rotation, having the required radius. Or, alternatively, a film might be used; but there is objection to the use of films because the shrinkage upon drying is not sufficiently uniform. In practice it is usually necessary to photograph only a small portion of the spectrum at one time and a short photographic plate is used without bending; the lack of fulfilment of the focusing condition is not enough to be harmful, and a small correction takes care of the difference which exists between the tangent of a small angle and the angle itself.

An accurately divided circle which serves to measure the basic angles is an important feature of spectrographs for soft X-rays.

It is to be noted that the above-described "focusing" does not lead to an increase in the intensity of the photographed spectrum. The length of ex-

posure which is required for photographing a given spectrum while rotating the crystal is equal to the sum of the exposure times required to cover the spectrum in a number of discrete contiguous sections, without rotating the crystal during each exposure. The primary advantage of rotating the crystal is to avoid the falsifications which were spoken of above.

A clock mechanism is frequently provided for rotating the crystal continuously, but for some types of work it suffices to rotate the crystal slightly at intervals during the exposure, by turning a slow-motion screw by hand.

The theory of the focusing condition was first given by W. H. and W. L. Bragg.* However, these observers explored the spectrum by the ionization method instead of photographically. In this method a metal chamber some 15 cm long and 5 cm in diameter is placed with its axis along the reflected X-ray beam. The chamber is provided at one end with a window, perhaps 1 cm square, covered with aluminum foil which is thin enough to be almost transparent to the X-rays. There is a slit in front of the window which is made narrow when measurements are in progress. The chamber contains a gas which is copiously ionized by the X-rays—for example, methyl bromide or argon. An electrode is provided within the chamber, and the wall of the chamber serves as the second electrode. A potential difference is initially applied to the electrodes and an electric current flows whenever an X-ray beam is entering the chamber—owing to the ionization caused by the X-rays. The magnitude of the ionization current is a direct measure of the intensity of the beam entering the slit. The strength of the current is measured by observing the number of scale divisions through which a sensitive electrometer is deflected in an arbitrarily chosen number of seconds of time. W. H. and W. L. Bragg, in their apparatus, fulfilled the focusing condition, and rotated the crystal in small steps, and each time rotated the ionization chamber through double the angle through which the crystal was rotated. Figure 201 shows one of the curves obtained; glancing angle is plotted as abscissa and intensity of ionization is plotted as ordinate. The target of the tube was of platinum and the reflection was from a cube face of rocksalt. The ionization maxima lettered A_1, B_1, and C_1 represent the first-order Bragg reflection of the group of spectrum lines which we now designate respectively as Pt $L\alpha$, Pt $L\beta$, and

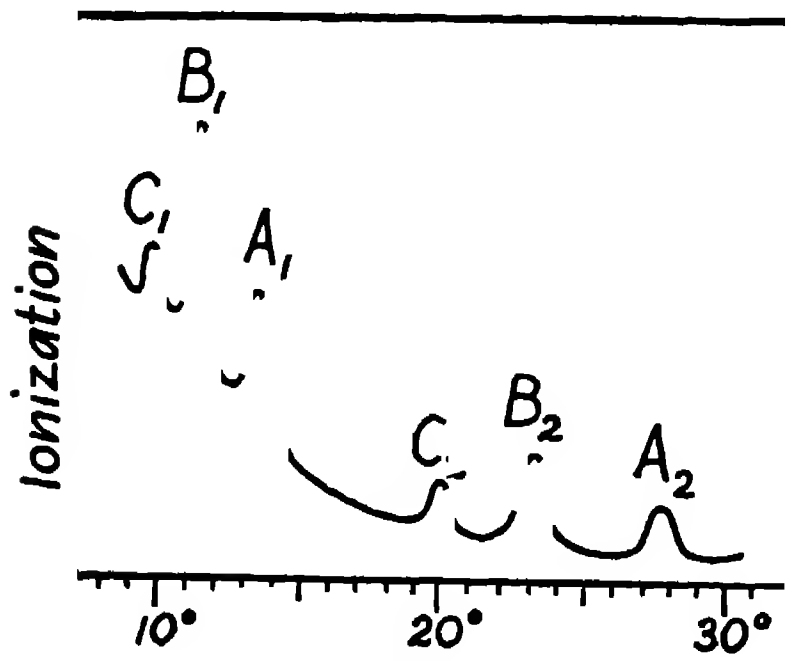


FIG. 201

* *Proc. Roy. Soc.*, **88**, 428, 1913.

Pt $L\gamma$. The individual lines of each group were not resolved in this early work. The maxima lettered A_2, B_2, and C_2 represent the second-order diffraction of these same groups of spectrum lines.

The ionization method, while more laborious than photography, is still much used; it has the advantage of yielding a direct measurement of the intensity which prevails at each point of the spectrum, and is moreover adapted to use in connection with the double crystal spectrometer, with which photography is not applicable.

The fulfilment of the focusing condition was first applied in conjunction with continuous rotation of the crystal and the photography of X-ray spectra by M. de Broglie.*

The most widely used modern design of the spectrograph for soft X-rays is that of M. Siegbahn.

For a more extensive account of various types of X-ray spectrographs than can be here given, see the well-known book by Siegbahn.†

The method of diffracting or reflecting from the *surface* of the crystal, which we have thus far alone considered, leads to accurate measurements of the positions of spectrum lines even without using the best of crystals. But the degree of perfection of the crystal affects intensity measurements materially. Crystals exhibit so-called "mosaic structure" to a greater or less degree. That is, a large crystal is, in general, made up of a multitude of microcrystals oriented just a little differently from each other. As a result, when the method of surface reflection is applied, the faces of the microcrystals do not all pass accurately through the axis of rotation located in the face of the entire crystal. But the amount by which they miss passing through this axis is small, and not sufficient to affect the focusing materially. Hence position measurements are not affected. But, for reasons which will appear later, the degree of disorientation which obtains does affect the intensity of the diffracted beam. In intensity measurements it is also a matter of importance whether the face be a natural one, or a cleavage face, or one which has been prepared by grinding and polishing parallel to a given set of net planes. When the face is a prepared one, the closeness of parallelism of the face to the net planes is an important factor in determining the intensity and sharpness of the spectrum lines.

Recently a type of spectrograph for soft X-rays has been developed which is based upon a principle advanced some twenty years ago but not until now brought to a stage of practical utility. The principle in question is to

* *Compt. rend.*, **157**, 924, 1913.

† M. Siegbahn, *Spektroskopie der Roentgenstrahlen* (2d ed.; Berlin: J. Springer, 1931), pp. 88 ff., or, in English, the translation of the first edition by G. A. Lindsay (Oxford University Press, 1925), pp. 50 ff.

bend a crystal of mica, for example, to the form of a cylinder of appropriate radius of curvature. The crystal will then perform in some respects like a concave grating, but in other respects quite differently. Suppose, with a curved crystal, a slit is placed upon the Rowland circle. An X-ray spectrum line of wave-length λ will be formed at the position which corresponds to the central image of the concave grating provided the slit happens to occupy the proper position on the circle such that the wave-length λ will undergo Bragg reflection. And there will be no other spectrum lines formed, either on the Rowland circle or elsewhere. Now let us suppose that the slit is dispensed with and that the target of the tube is placed with its face along the Rowland circle. A beam will diverge from each point of that portion of the Rowland circle along which the bombarded portion of the target extends. Hence a range of wave-lengths will be recorded along the corresponding portion of the Rowland circle on the opposite side of the normal to the crystal face. But the effective width of the crystal must not be too great, lest excessive aberration result. Each wave-length will be reflected from the entire effective portion of the crystal face *simultaneously*, hence there will be a great gain in intensity as compared to the plane crystal. Furthermore, there is really no need for the face of the target to lie along the Rowland circle. The target may also be placed quite close to the crystal and then a greater range of wave-lengths may be photographed at one setting.

Alternatively, a slit may actually be placed upon the Rowland circle, and backed by the target as in the plane crystal spectrograph. Then a sliding motion of the crystal and photographic plate or film along the Rowland circle is necessary during exposure, however. Another variant of the curved crystal spectrograph resembles some of the plane crystal spectrographs for hard X-rays to be described in the next section—in so far as the radiation passes *through* the crystal.

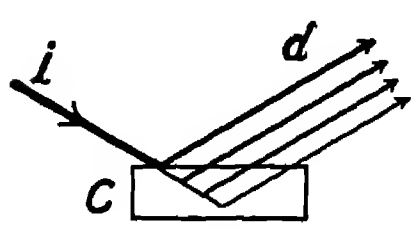


FIG. 202

For further information concerning curved crystal spectrographs, see the references given below.*

8. Volume Effect; Spectrographs for Hard X-rays.—When a spectrograph of the type shown in Fig. 200*a* is used with hard X-rays, each incident ray penetrates into the crystal to a considerable depth, as represented in Figure 202, and hence the diffracted beam will be broad, and the spectrum lines will appear broad. Moreover, the spectrum lines will be shaded off to one side because the incident ray diminishes in intensity along its path in the crystal. Hence the intensity of the beam which is diffracted will

* H. H. Johann, *Zeit. f. Phys.*, **69**, 185, 1931; Y. Cauchois, *Jour. de phys.*, **3**, 320, 1932; L. von Hamos, *Annalen*, **17**, 716, 1933; T. Johannson, *Zeit. f. Phys.*, **82**, 510, 1933; A. Sandstrom, *ibid.*, **84**, 541, 1933.

diminish progressively; moreover, the less intense rays diffracted at the lower levels will themselves undergo a diminution of intensity before emerging from the crystal. The effect which is here in question is called the "volume effect."

The broadening and shading off of the spectrum lines is of course objectionable. It causes both a lowering of resolving power and a lowering of the possible accuracy of measurement of the positions of the spectrum lines.

H. Seemann* has devised two spectrographs used with hard X-rays which we shall describe. One of these, represented in Figure 203, is known as the "wedge spectrograph." The rays originate at the target, T, and are incident upon the crystal, C, under a wedge, W, and are then "reflected" to the

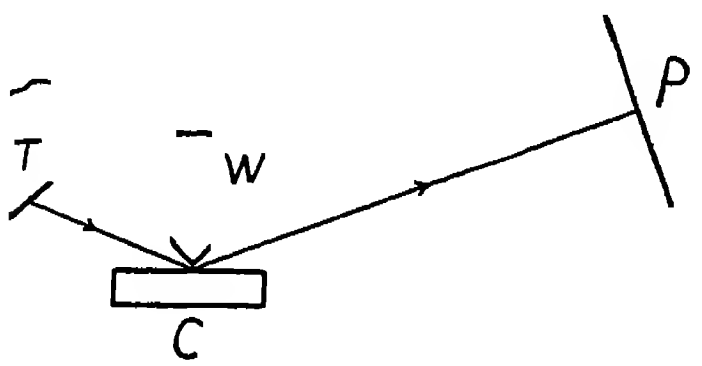


FIG. 203.—Seemann wedge spectrograph

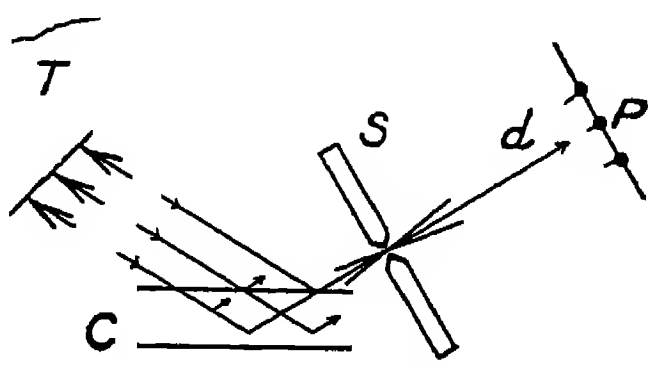


FIG. 204.—Seemann slit spectrograph

photographic plate, P. The wedge is not quite in contact with the crystal; there is a small amount of clearance. The effective width of the slit which is thus formed is equal to the distance of the wedge from its mirror image in the surface of the crystal.

Since the target, T, is extended in area, a range of glancing angles occurs and a small region of the spectrum may be photographed at a single setting. But commonly the crystal, wedge, and photographic plate are rotated as a unit, relatively to the target, in order to photograph a larger region of the spectrum and avoid danger of introducing falsifications into the spectrum arising from points of the target which give especially intense radiation. This spectrograph is simple and convenient. It is suited for work with hard X-rays in so far as it is suited for work at small glancing angles, and also because the spectrum lines obtained with it show the features caused by the penetration of the beam into the crystal to a less degree than do the spectrum lines obtained with a spectrograph designed for soft X-rays, but they are not free from these features.

A spectrograph subsequently devised by Seemann is represented in Figure 204. The rays diverge from various points of the bombarded area of the target, T, and are incident upon the crystal, C, and then pass through the slit S. The crystal must be a good one and fairly large, but because of

* *Annalen*, **49**, 470, 1916, and *Phys. Zeit.*, **18**, 242, 1917.

the penetrating quality of the rays in question, the nature of the surface of the crystal now has little influence. Out of all of the rays from various points of the target which are incident upon the crystal at the same glancing angle, the same wave-length is reflected. The reflected beams give rise to the ray, d, which, passing through S, falls upon the photographic plate at P. Owing to the divergence of the beams leaving the target, a range of glancing angles obtains, and a small region of the spectrum may be photographed at one setting. But in order to photograph a wider spectral region, the crystal, slit, and plate-holder are usually rotated as a unit, in relation to the target. This spectrograph is well suited for accurate work with hard X-rays, for the resulting spectrum lines do not exhibit the features due to volume effect. The reason the instrument avoids these features is because of the location of the crystal in *front* of the slit. If the direction of the beam were reversed, the rays originating along the line P and passing through the slit before striking the crystal, the photographic plate now being along the face of T, it is obvious that broadening and shading-off of the spectrum lines would then occur.

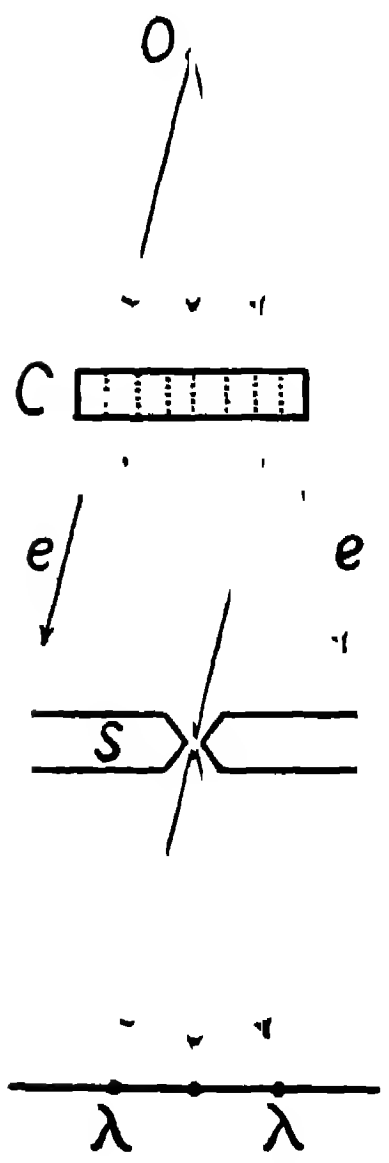


FIG. 205

Another type of spectrograph suited for work with hard X-rays was originated by Rutherford and Andrade,* who, however, used it only for investigation of the γ-radiation from Ra B and Ra C, but it is suited for work with hard X-rays as well. It is represented in Figure 205. A beam from an X-ray tube is incident from above in the figure, upon a first slit located at O. The beam which diverges from O falls upon a crystal, C, 1 or 2 mm thick, and here the rays are "reflected internally" by the net planes indicated by the dotted lines. The glancing angle which any incident ray makes with this set of net planes is a function of the distance to the right or left of the center of the crystal at which the ray is incident. Corresponding to the different glancing angles which the different rays make with the net planes, there are different wave-lengths reflected. The diffracted beams will, however, all intersect at a distance CS beyond the crystal equal to the distance OC. Accordingly, a second slit S may be placed at this point. A photographic plate is then placed at any suitable distance beyond the second slit and the spectrum is formed upon it. Each spectrum line appears symmetrically at equal distances to the right and left of the central image.

The spectrograph of Rutherford and Andrade was improved in design

* *Phil. Mag.*, **28**, 263, 1914.

by J. M. Cork,* who used it in its improved form for making many excellent wave-length determinations in the hard X-ray region.

Still other spectrographs suited for the study of hard X-rays are described in the book by Siegbahn referred to above.

Rutherford and Andrade originally used their spectrograph without the second slit S, simply placing the photographic plate at some convenient distance beyond the point S where all of the diffracted beams intersect. When the spectrograph is so used, a phenomenon of significance is strikingly revealed. Obviously the same spectrum will be formed whether the slit S be present or not, but when this slit is absent there will be superposed upon the spectrum the illumination which arises from that fraction of the beam issuing from the slit O which is not absorbed by the crystal and passes through it directly, that is, without diffraction. Now it is found that this directly transmitted illumination is not uniform across the photographic plate, but shows discontinuities as follows: Let λ be a wave-length which gives rise to dark spectrum lines on the photographic negative at the positions indicated by λ in our figure. It is found that *light* lines are recorded at the points where the directions e intersect the photographic plate, that is, the recorded spectrum now consists of corresponding pairs of dark lines and light lines. The occurrence of these light lines proves that when an X-ray beam of wave-length λ is incident upon a given set of net planes of the crystal at the proper angle to be diffracted according to Bragg's law, and is diffracted, it is extinguished with markedly greater rapidity in its passage through the crystal than when it is incident at some other angle, at which it is not diffracted. The fact that more rapid extinction occurs when a beam is being diffracted was also proved in an especially designed experiment performed by W. H. Bragg.†

If actual crystals were "ideal," instead of consisting of microcrystals slightly disoriented, the rate at which the direct X-ray beam is extinguished within the crystal when the beam is being diffracted would be even greater than it is. The slight disorientation of the microcrystals means that only a fairly small fraction of these may have the proper orientation relative to any given direction of incidence, to take part in the diffraction. And within those microcrystals which do not take part in the diffraction, the rate of extinction of the beam is only that determined by the ordinary mass absorption of the material constituting the crystal, as determined by its density. Hence if all of the microcrystals were correctly oriented, the extinction would be more rapid than it is. According to theoretical treatments developed by Darwin, Ewald, Prins, and others, a beam which is under-

* *Phys. Rev.*, **25**, 197, 1925; see also Cork and Stephenson, *ibid.*, **26**, 138 and 530, 1926.

† *Phil. Mag.*, **27**, 881, 1914.

going diffraction would penetrate an ideal crystal to a depth of only a few thousandths of a millimeter, even if the beam were one of hard X-rays.*

9. Methods of Crystal Analysis.—Having considered some of the instruments used in recording X-ray spectra, let us now consider some of the instruments which are employed in determining the arrangement of the atoms within crystals.

For further information than can be here given, see the references in the footnote.†

Laue's discovery of the diffraction of X-rays by crystals was followed, in the same year, by the promulgation by W. L. Bragg of his view, which regards diffraction as a reflection. Thereupon followed in the years 1913 and 1914 numerous further investigations of the diffraction of X-rays by crystals; as an example of these we may take the classic researches of W. H. and W. L. Bragg carried out by the ionization method in which one of the curves obtained was that of our Figure 201. From an experimental standpoint it was not safe for these investigators to conclude at once that the maxima of this curve represent spectrum lines characteristic of the material of the target, which was platinum in this case. With little positively known at the time, the maxima might have been in some manner characteristic of the crystal, which was rocksalt. But by varying the material of the target, a different set of maxima was obtained each time, hence they could conclude that the maxima do represent the spectrum of the target. The detail of the structure of the rocksalt crystal, and of other crystals, was, however, unknown, hence the D of Bragg's law could not at once be determined, and hence the wave-lengths could not be determined at once even after the glancing angle of reflection and the spectral order had been ascertained. However, from Avogadro's number, which was known, and the known density of the crystal, the various possible sizes of the unit cell of the crystal could be calculated upon the assumption that the unit cell contains one molecule per unit cell, or two molecules per unit cell, or three, or four, etc., the number being perforce integral. But by comparing the various positions and intensities of a given maximum due to a given target material

* C. G. Darwin, *Phil. Mag.*, **27**, 315 and 675, 1914, **43**, 800, 1922; P. P. Ewald, *Phys. Zeit.*, **26**, 29, 1925; J. A. Prins, *Zeit. f. Phys.*, **63**, 477, 1930.

For summaries of the theory which treats of the rate of extinction both within ideal crystals and within crystals which are not ideal, see A. H. Compton, *X-rays and Electrons* (New York: Van Nostrand & Co., 1926), chap. 5, and E. O. Wollan, *Rev. Mod. Phys.*, **4**, 205, 1932.

† W. L. Bragg, *An Introduction to Crystal Analysis* (London: Bell, 1928); R. W. G. Wyckoff, *The Structure of Crystals* (New York: Chem. Catalog Co., 1931); Geiger and Scheel, *Handb. d. Physik*, Vol. **24**; and A. Schleede and E. Schneider, *Roentgenspektroskopie u. Kristallstrukturanalyse* (Berlin and Leipzig: De Gruyter & Co., 1929), in two volumes.

when the reflection was first from a (1, 0, 0) face of rocksalt, then from a (1, 1, 1) face, and then from a (1, 0, 1) face, etc., and doing the same with other crystals, it was possible to work out the arrangement of the atoms within simple crystals with practical, but not absolute, certainty. The arrangement for rocksalt which was arrived at was that of our Figure 194, comprising four molecules per unit cell. In the figure the length of side of the unit cell is indicated by a. With this arrangement established, X-ray wave-lengths could be calculated. Moreover, the wave-lengths calculated on this basis were in accord with those to be expected according to the fundamental quantum relation, equation (1). This agreement confirmed the analysis which had been made of the arrangement of the atoms in rocksalt and other simple crystals.

In view of the foregoing, it should be evident that the instruments which are used for recording X-ray spectra are also instruments of crystal analysis. Investigation by these instruments, considered as a group, is thus to be regarded as one method of attacking the problems of crystal structure. The spectrograph of the type now primarily used for soft X-rays, especially, has yielded important results concerning the arrangement of the atoms within crystals.

The Laue method of diffracting X-rays by crystals, of which we have not yet given the theory, is very valuable in ascertaining the symmetry of a crystal, and is to be regarded as the second of the principal methods employed in crystal analysis, of which there are four. The Laue method is generally used in conjunction with one or more other methods.

10. Rotation Photographs.—One of the two remaining methods of crystal analysis will be explained in referring to Figure 206. Let i be a narrow horizontal beam of X-rays of wave-length λ, the pencil of rays being limited by a succession of small circular apertures, or, under some circumstances, by vertical slits, and let C be a small crystal. Ideally, the beam should be monochromatic, but in practice it is sufficient if the spectrum of the beam be very simple, consisting, for example, of one dominant spectrum line and two or three relatively faint ones. Such a spectrum may be obtained by proper adjustment of the voltage actuating the tube and, if necessary, absorbing undesired harder radiation by means of filters.

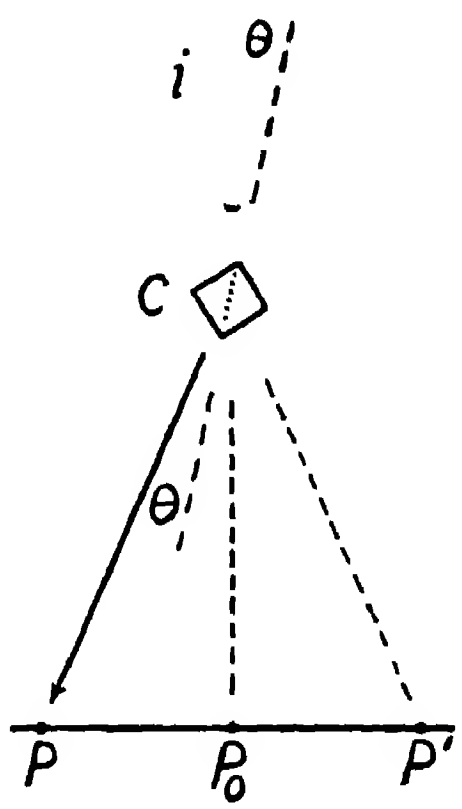


FIG. 206

The beam should be hard enough to penetrate the small crystal readily, as the method in question depends upon the occurrence of internal reflections.

The crystal is mounted with an important crystallographic direction perpendicular to the plane of the figure, or vertical, and is then rotated about this direction as axis. The rotation may be either continuous or may take the form of an oscillation. Let the dotted line within C be parallel to an important set of net planes. Each time, in the course of the rotation, that this set of net planes makes the proper glancing angle, θ, with respect to the incident beam, the wave-length λ will be reflected to the point P of the photographic plate, PP', upon which the reflection will be recorded. As the crystal rotates there will also be a corresponding spot formed at P', on the other side of the point P_0 where the undeviated beam strikes the plate. The value of the angle θ is now determined from the relation $\tan 2\theta = PP_0/P_0C$, and, since λ is known and m is readily ascertainable, the separation, D, of the reflecting net planes can be calculated. Other sets of net planes give rise to other pairs of spots, and by measuring the distance corresponding to PP_0 for each spot, the separations of the various sets of net planes can be determined. From this knowledge, when complete for all sets of net planes having physical importance, the arrangement of the atoms within the crystal may be deduced. The process is a long and difficult one, however, especially in the case of complicated structures. We shall not attempt to enter into it. Due regard must be paid to the intensities of the spots; the intensities are sometimes merely estimated and sometimes measured photometrically.

The spots which occur along the line PP' of our figure originate from net planes which are parallel to the axis of rotation of the crystal. But when sets of planes which are inclined to this axis, in the course of the rotation make the proper angle θ with the incident beam, these planes will also give rise to reflected beams—and these beams will form spots which occur on the photographic plate above and below the plane of the diagram. The pattern of spots which is thus formed is symmetrical about a vertical line through P_0 and is also symmetrical about the horizontal line PP', which is spoken of as the "equator" of the pattern. The spots lie along the intersections of a system of nearly straight horizontal lines which are portions of flat hyperbolas, with a system of less flat hyperbolas which run in the general vertical direction.

Each important crystallographic direction is used in turn as the axis of rotation of the crystal.

In place of the photographic plate a film is sometimes used, which is then bent around the crystal at a radius equal, let us say, to CP_0, so that it forms a circular cylinder of which the axis coincides with the axis of rotation of the crystal.

The method of rotation photographs, as applied specifically to the analysis of crystals, was first described by E. Schiebold in the year 1919

and was further developed by him and by Polanyi and Weissenberg in 1921. But only from about 1924 on has the value of this method been widely appreciated.*

11. The Powder Spectrograph.—The fourth principal method of crystal analysis will be explained in referring to Figure 207. Let i be a narrow horizontal beam of X-rays of wave-length λ. For the purpose of preliminary discussion let this beam be considered as limited always by a succession of minute circular apertures. Let C be a small quantity of a crystalline compound in the form of a fine powder. Consider the beam as monochromatic—though in practice the spectrum may consist of a dominant spectrum line accompanied by one or two relatively faint ones.

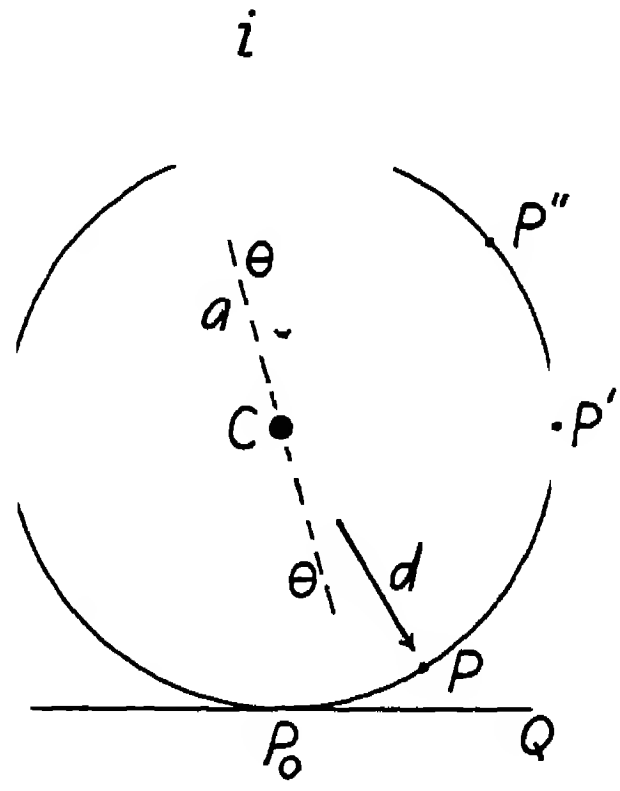


FIG. 207

The minute crystals of the powder will be oriented in all directions, practically speaking. Hence there will be some crystals which are properly oriented to reflect the wave-length λ from each important set of net planes. Let the broken line a indicate the orientation of a set of net planes in a given minute crystal; the planes in question are supposedly perpendicular to the plane of the figure. They make an angle θ with the direction of the incident beam, and if the angle θ is such that the relation $m\lambda = 2D \sin \theta$ is satisfied, there will result a diffracted beam d. Suppose the angle θ is such that the diffracted beam is formed. There will also be other minute crystals in which the same set of net planes makes the same glancing angle θ with the incident beam, but these net planes will not always be perpendicular to the plane of our figure. But since θ is always the same for the given set of net planes in the various crystals under consideration, the direction of diffraction or reflection will generate a circular cone about the direction of incidence as axis.

For each other set of net planes, of different separation, there will be a circular cone of different angle. Hence if a photographic plate, P_0Q, be placed perpendicular to the incident beam, there will be formed upon this a series of circles about P_0 as center, one circle for each cone. These circles are often spoken of as "spectrum lines," but we must remember that they are all due to the same wave-length λ, only reflected in different sets of net planes.

* See Polanyi, Schiebold, and Weissenberg, *Zeit. f. Phys.*, **23**, 337, 1924; E. Schiebold, *ibid.*, **28**, 355, 1924; and J. D. Bernal, *Proc. Roy. Soc.*, **113**, 117, 1927, where a complete description of the method itself is also to be found.

In place of the plate P_0Q, a film P_0P'' may surround the powdered material. The "spectrum lines" will then be roughly circular arcs about the direction iP_0 as axis. However, if a line happens to be located at P', this line will be perfectly straight. When the film is laid out flat it will appear as represented in Figure 208. The letters under this figure correspond to the letters of Figure 207. It is to be noted that the lines on opposite sides of P' are curved in opposite directions.

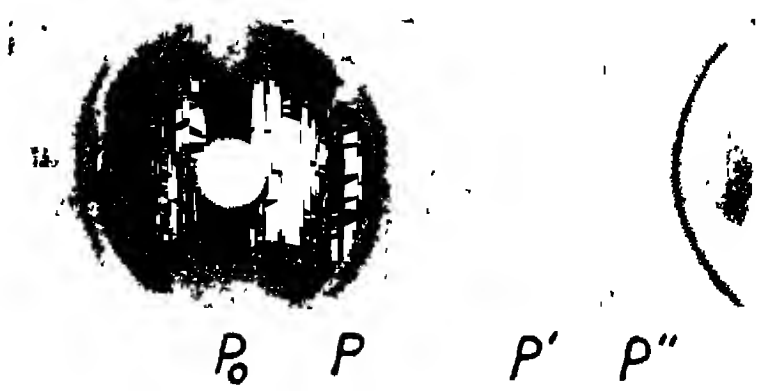


FIG. 208

The distance of each spectrum line from the point P_0 is measured, and thus the angle θ which corresponds to each line, and hence the separation of the corresponding net planes, is ascertained. The separations of the net planes, considered in conjunction with the intensities of the lines to which they give rise, furnish the key for determining the arrangement of the atoms within the crystal.

The powder method is of great value, especially because it is applicable to compounds which can be obtained only in minute crystals. It can also be applied, for example, to a polycrystalline metal foil or sheet, in which the microcrystals are oriented more or less at random. Further, when a foil or sheet has been formed by rolling, the microcrystals usually show a preferred orientation, with respect to the direction of rolling. This fact is then revealed by an increased intensity of the spectrum lines along two diametrically opposite portions of their contour and a lessened intensity along a diameter perpendicular to the first. When a few unusually large microcrystals are present in the material, some of these may give rise to isolated spots of especially great intensity along the spectrum lines.

Before the day of the powder spectrograph, any substance was considered to be amorphous when examination under a good microscope failed to reveal crystal structure. But it has now been found that many substances which were adjudged amorphous by this criterion prove to be crystalline when they are subjected to examination in the powder spectrograph—and it has been possible in a multitude of cases to determine the crystal structure. However, when sufficiently large crystals are obtainable, the method previously described, of rotating a single crystal about one known crystallographic direction at a time, is more powerful as a means of unraveling the detail of complicated structures.

Since a powder spectrogram need be measured only in the horizontal direction, and not diagonally also, like a rotation photograph, the incident beam may be delimited by vertical slits in place of minute circular apertures; since the use of slits presents obvious advantages, it is common practice to use them in the powder spectrograph.

The powder spectrograph was originated by Debye and Scherer* in the year 1916 and independently by A. W. Hull† in 1917; it came into general use some six or eight years before the method of rotation photographs.

12. The Laue Theory (*Introductory*).—The reader will recall that in the experiment of Laue, Friedrich, and Knipping, described in section 3, a beam of X-rays is limited by a succession of circular apertures about a millimeter in diameter, and is then incident upon a crystal. The beams diffracted by the crystal are received upon a photographic plate. The incident beam is heterogeneous in wave-length, or, in other words, consists of what is often called "white radiation." We shall now present the theory which led Laue to propose this celebrated experiment.

In the theory, the effect of a space lattice upon one given wave-length at a time is considered. That is, we are to solve the problem first for a monochromatic beam of wave-length λ.

For the sake of simplicity, we shall suppose the lattice to be simple cubic.

For reasons already explained in section 5, we may consider the incident waves as being plane so long as we concern ourselves with only a small portion of the crystal at one time, and the diffracted waves are then also to be considered as being plane. If we were dealing with *light*, the incident beam might be made accurately parallel by a lens, and parallel diffracted rays might be brought to a focus by a second lens. For a point source there would then result upon the photographic plate, just as with a grating, a point image in monochromatic light, or a spectrum *dot*. With X-rays, since we have no lenses, the incident beam is merely restricted in cross-section by screens with small circular apertures, and is thus slightly divergent. As a result of this slight divergence and the failure to focus the diffracted rays, what would be a spectrum-dot is spread into a patch or *spot*. The size of this spot depends upon the diameter of the apertures and upon the distance between these, and also upon the position of the crystal and of the photographic plate relatively to the apertures. However, the theory does not concern itself immediately with the size of the spot, which is determined by the divergence of the beam considered as a whole. It is sufficient that for any elementary portion of the beam, of cross-section of let us say .01 mm^2, the incident and diffracted waves may be regarded as plane.

We shall limit the discussion here to the case in which the beam is incident upon the lattice in the direction of the positive Z-axis (see Fig. 193); the beam is to be supposed as incident from the rear. The more general case, of any direction of incidence, is treated in Appendix G.

The origin of co-ordinates is to be taken as indicated, at the center of one of the lattice particles. The lattice supposedly extends along the negative

* *Phys. Zeit.*, **17**, 277, 1916. † *Phys. Rev.*, **10**, 661, 1917.

axes as well as along the positive ones. Our problem is to determine in what directions, and for what wave-lengths, the Huygens wavelets originating from the various points of the lattice will all reinforce each other. The method of procedure which we shall adopt is to determine in turn the nature of the diffraction pattern which is produced by each of the following systems:

1. By a single row of lattice particles perpendicular to the incident beam. The pattern arising from the row of particles lying along the X-axis will first be deduced, and then the pattern arising from the row of lattice particles lying along the Y-axis, which will be similar to the first, only rotated through 90°.

2. By the array of lattice particles in the XY-plane. Such an array is called a "regular plane net," or simply a "net." The net in question will be considered as composed of horizontal and vertical rows. The diffraction from crossed gratings, which is similar, will be considered in this connection.

3. By the row of lattice particles lying along the Z-axis.

4. Finally, we shall consider the diffraction due to the entire space lattice, viewing the lattice as composed of a succession of "nets" equally spaced along the Z-axis.

13. Diffraction by a Row of Particles.—Referring to Figure 209, which is drawn in the XZ-plane, let the circles represent the row of lattice particles

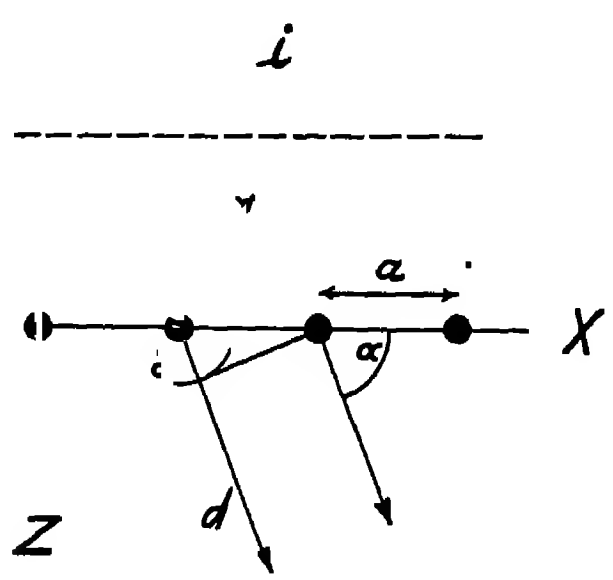


FIG. 209

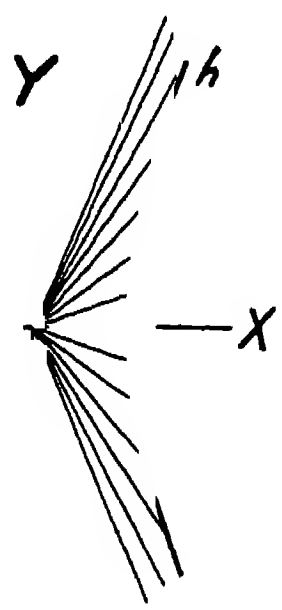


FIG. 210

lying along the X-axis. Each incident wave front reaches all of the particles simultaneously. The Huygens wavelets originating from these particles will reinforce each other in any direction, d, in which the path difference δ is an integral number of wave-lengths. Now δ equals $a \cos \alpha$, where a is the lattice constant and α is the angle made by the direction d with the X-axis. Hence we have, as the condition for reinforcement:

$$m_x\lambda = a \cos \alpha , \tag{4}$$

where m_x is the order of diffraction, which may take any integral value, including zero. The direction of diffraction, d, for any given order, m_x, may be rotated about the X-axis without changing the value of δ, consequently the direction of diffraction, for any given order, m_x, generates a circular cone making an angle α with the X-axis. This cone should be thought of as one of wide angle. It is represented in Figure 210 as it would be seen by an observer who is looking toward the incident beam of X-rays; the irradiated row of particles is supposedly comprised entirely within the short horizontal line segment which penetrates the apex of the cone. The intersection of the cone with a plane perpendicular to the Z-axis will form one branch of a hyperbola h. Each spectral order will give rise to its own hyperbola. In Figure 211 the branches for the first and second orders are represented by

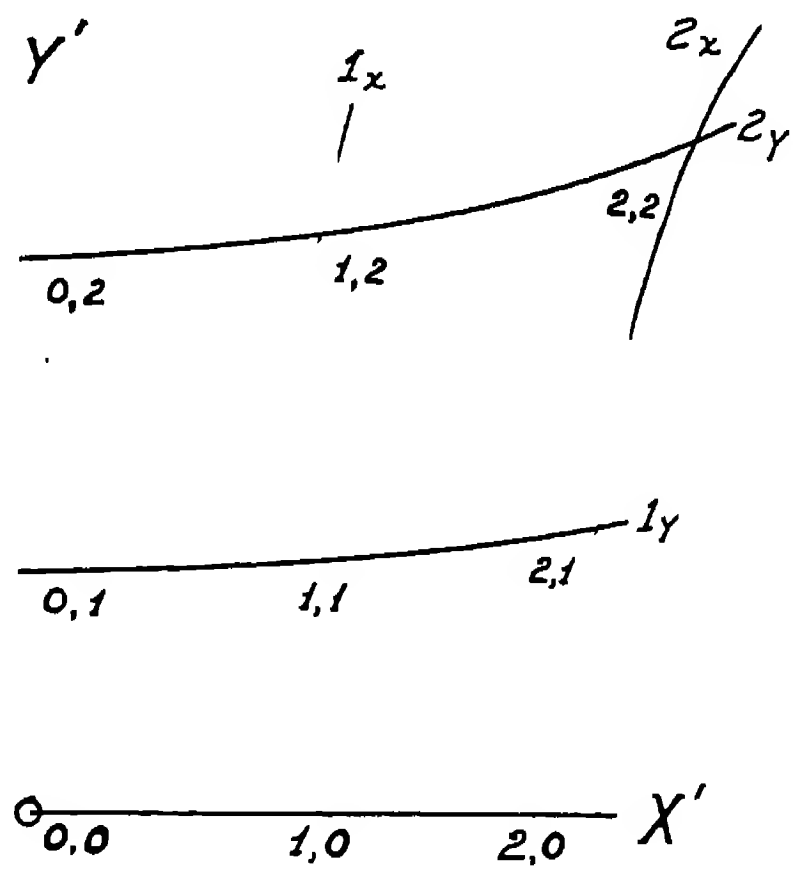


FIG. 211

1_x and 2_x, respectively. The hyperbola of zero order is a vertical line, namely, the Y'-axis of this figure. The branches for the negative orders fall to the left of the Y'-axis, the value of α being greater than 90° for these orders. If it were possible to record the diffraction pattern for a single horizontal row of lattice particles, the principal maxima, due to one wave-length λ, received upon a photographic plate erected perpendicularly to the Z-axis, would consist of a family of hyperbolas running in the general vertical direction.

It is to be noted that in passing from one lattice particle to the next in the positive X-direction, the length of optical path by way of successive particles decreases for the positive orders, and increases for the negative orders.

Considering now the diffraction due to the row of particles lying along the Y-axis: The direction of diffraction, for a given order, to be denoted by m_y, generates a circular cone making an angle β with the Y-axis, the angle β being determined by the equation:

$$m_y\lambda = a \cos \beta \,. \tag{4'}$$

This cone intersects a plane perpendicular to the Z-axis in one branch of a hyperbola running in a general horizontal direction. The various orders of diffraction give rise to a family of hyperbolas running horizontally. The branches for the first and second orders are indicated in Figure 211 by the curves 1_y and 2_y, respectively.

14. Diffraction by a Plane Net.—Figure 213 represents the net of particles lying in the XY-plane. The beam is incident perpendicularly from

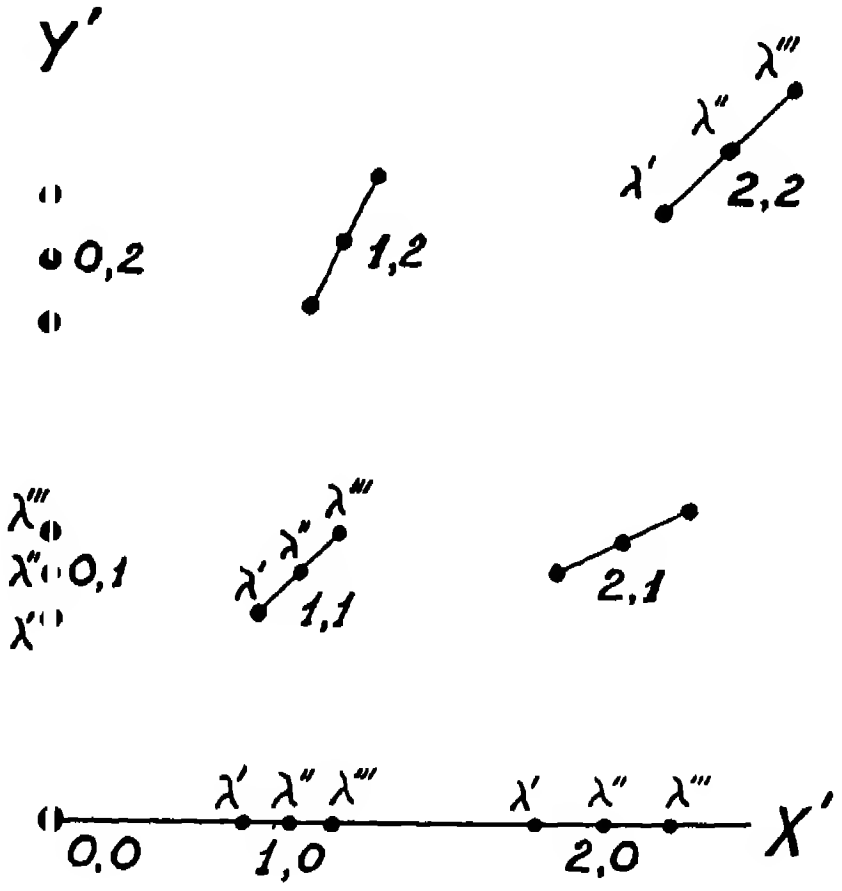


FIG. 212

the rear. The condition for the formation of a principal maximum of diffraction by the net is that the Huygens wavelets from all lattice points in the array shall reinforce each other. In order that this may happen, the wavelets from each horizontal row must reinforce each other and, simultaneously, the wavelets from each vertical row must reinforce each other. In other words, equations (4) and (4′) must be simultaneously satisfied. These equations are both satisfied along the intersection of any one of the cones of angle α, given by equation (4), with one of the cones of angle β given by equation (4′). Now the intersection of any two of these cones, one of each family, penetrates a plane perpendicular to the Z-axis at the intersection of the two branches of the hyperbolas, one of each family, which

correspond to the cones (see Fig. 211 again.) In this figure the various intersections of the hyperbolas of the two families are indicated by double indices. The first number of the index denotes the order with respect to the X-axis, and the second number indicates the order with respect to the Y-axis. These intersections represent the positions of the various principal maxima of diffraction due to a single plane net, for one given wave-length λ. The irradiated portion of the lattice projected into this figure would be comprised in a small circle such as that indicated around the origin of co-ordinates.

Let us now suppose that the incident radiation contains three wave-lengths instead of a single one, namely, λ', λ'', and λ''', such that $\lambda' < \lambda'' < \lambda'''$. There will then result the array of spectrum dots shown in Figure 212. The spectral order of each set of three dots is indicated as in Figure 211, by double indices. In each set the dot corresponding to λ' lies nearest the origin of co-ordinates and the dot corresponding to λ''' lies farthest away. If instead the spectrum of the incident radiation were continuous, extending over a given limited range, then each set of spectrum dots would be replaced by a line segment, which, if we may employ optical terminology, would be colored violet at the end nearer the origin of co-ordinates and red at the farther end.

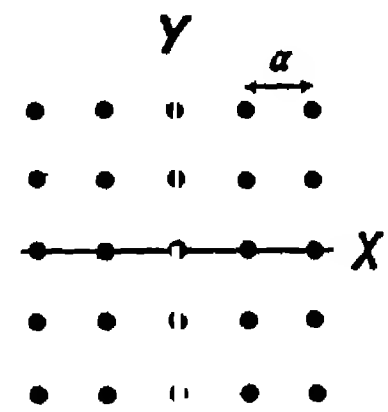


FIG. 213

15. Crossed Gratings.—The same laws which govern the formation of spectra by a plane net also govern the formation of optical spectra by crossed diffraction gratings. Consider two transmission gratings consisting each of wide opaque strips and narrow transparent lines, both having the same grating space a. When these gratings are placed in contact, with their lines crossed at right angles, they form a rectangular array of tiny apertures. Each aperture acts as a source for transmitted Huygens wavelets. The array of apertures is equivalent to a square array of lattice particles spaced at the distance a. The apertures, however, send no wavelets in the backward direction, as do the lattice points, so the equivalence holds only in the region of "transmission."

If the two gratings were separated, their planes remaining parallel, and their lines remaining crossed at right angles, the diffraction effects due to the gratings would not alter. In Figure 214, which is supposedly horizontal, O is a point source of light; L_1 is a collimating lens; A is the first grating, with lines vertical; B is the second grating, with lines horizontal; L_2 is a focusing lens; and P is a photographic plate. The incident beam passes in part through both gratings without deviation, forming a central image at C. The incident beam is also in part diffracted horizontally by the first

grating, and the beams thus diffracted pass in part through the second grating without deviation, resulting in the formation of horizontal spectra at h_1 and h_2 just as though the second grating were absent. Again, a portion of the beam which passes through the first grating without deviation is diffracted vertically by the second grating, resulting in the formation of vertical spectra above and below the plane of the diagram, at C. Moreover, a beam diffracted horizontally by the first grating is in part again diffracted by the second grating, resulting in the formation of diagonal spectra. The diffracted beam leaves the first grating at an angle α (see Fig. 215), this angle being determined by equation (4). Various points along any one transparent line of the second grating, B, act as the centers of Huygens wavelets. The beam must be diffracted by the second grating in such a direction that the phase relations existing at points along this line are maintained in the diffracted beam. Thus the direction of diffraction must lie on a cone of angle α about a horizontal axis. Furthermore, points lying vertically above each other in adjacent lines of grating B must have phase relations as given by equation (4′). The direction of diffraction must therefore lie also on the cone of angle β about a vertical line, or, in other words, it must lie along the intersection of the two cones.

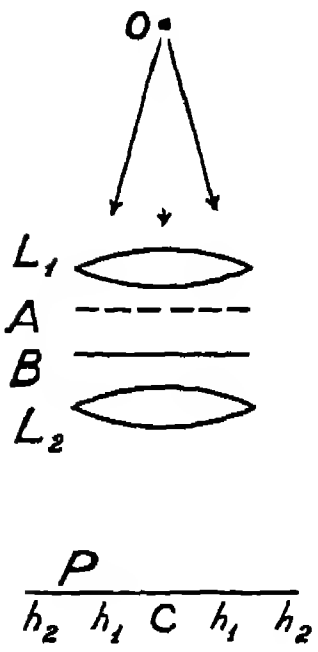


FIG. 214

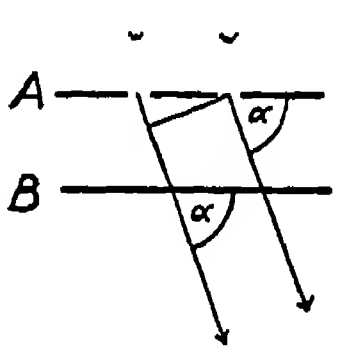


FIG. 215

In the present case, the apices of the intersecting cones must be considered to lie at the center of the lens L_2, because the one ray of each beam leaving the second grating which passes through the center of this lens, continues on its course without deviation. The size of the final pattern increases in direct proportion to the focal length of the lens L_2, just as it does for a single grating.

In the foregoing discussion it has been supposed that the gratings are formed of wide opaque strips and narrow transparent lines. The theory, however, applies also to other transmission gratings—for example, to "replicas"—and it applies also to reflection gratings when these are suitably mounted.

16. Diffraction by a Row of Particles (***Concluded***).—We now return to the consideration of the diffraction caused by a single row of lattice particles. This time we shall suppose that the row lies *along* the beam, and not perpendicular to it, as previously. The present case will be discussed in referring to Figure 216, in which the row of particles illustrated is the one lying along the Z-axis of co-ordinates. A given wave front now reaches the

particles *in succession*. The wave must travel the additional optical distance a before reaching each successive particle. Let us compare the optical paths from the origin O to the diffracted wave front ME, by way of particles *1* and *2*, respectively. The path by way of *2* exceeds that by way of *1* by an amount $(a-a\cos\gamma)$, where γ is the angle between the direction of diffraction which is under consideration and the Z-axis. This path difference must equal an integral number of wave-lengths if reinforcement is to take place. The fact that the path by way of particle *2*, with larger positive co-ordinate, *exceeds* that by way of *1*, with smaller positive co-ordinate, means that the order of diffraction with regard to the Z-axis, m_z, must be taken as negative. When the incident beam has the direction of the positive Z-axis, as here supposed, all values of m_z are negative.

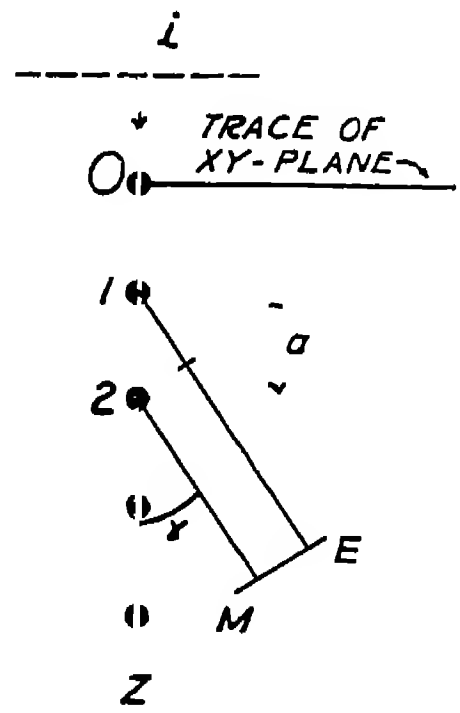


FIG. 216

In view of the foregoing, the condition for constructive interference is:

$$m_z\lambda = a\ (\cos\gamma - 1)\ . \tag{4''}$$

For each given order of diffraction, m_z, the direction of diffraction generates a circular cone about the Z-axis of which the elements make a constant angle γ with this axis. The cone for each order intersects a plane perpendicular to the Z-axis in a circle. The innermost circle is that of order -1, the next circle is that of order -2, etc. Thus, for each wave-length, λ, the principal maxima of diffraction for a row of lattice particles lying along the Z-axis, received upon a photographic plate perpendicular to this axis, form a family of circles about the Z-axis as center.

17. The Space Lattice.—We are now ready to consider the diffraction arising from an entire space lattice, viewing the lattice as a succession of "nets" equally spaced along the Z-axis.

The disturbances contributed by successive nets will reinforce each other in those directions and only in those directions in which the disturbances from a row of particles taken perpendicular to the nets will reinforce each other. That is, the disturbances from successive nets will reinforce each other along the cones of angle γ determined by equation (4″). However, each net considered separately sends an appreciable disturbance only in the directions of the intersections of the cones of angle α and β about the X- and Y-axes, respectively, determined by equations (4) and (4′). Whether there will be anything appreciable to reinforce from the successive nets along the cones of angle γ depends, then, upon whether a cone of angle α about

the X-axis, one of angle β about the Y-axis, and one of angle γ about the Z-axis, each with its apex at the origin of co-ordinates, all intersect in one line. If they do, there will be a principal maximum formed. If the three cones do not intersect in one line there will be no principal maximum formed by the space lattice as a whole considered as a diffracting system. In general, three cones about three axes do not intersect all in the same line. We must therefore conclude that there will in general be no principal maximum formed. Under what special conditions, then, may a principal maximum be expected? If α, β, and γ are all to apply to one and the same direction in space, the following geometrical relation must be satisfied:

$$\cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma = 1 . \tag{5}$$

Assembling equations (4), (4′), and (4″):

$$m_x \lambda = a \cos \alpha , \tag{4}$$

$$m_y \lambda = a \cos \beta , \tag{4′}$$

$$m_z \lambda = a (\cos \gamma - 1) . \tag{4″}$$

Substituting the values of the cosine functions from these three equations in equation (5), we obtain:

$$\left(\frac{m_x \lambda}{a}\right)^2 + \left(\frac{m_y \lambda}{a}\right)^2 + \left(\frac{m_z \lambda}{a}\right)^2 + \frac{2 m_z \lambda}{a} = 0 . \tag{6}$$

From this expression λ/a can be factored out and discarded as a solution, for neither $\lambda = 0$ nor $a = \infty$ has any interest. Solving the remaining expression for λ, we have:

$$\lambda = -\frac{2 m_z a}{m_x^2 + m_y^2 + m_z^2} . \tag{7}$$

The wave-length, λ, must satisfy this equation if there are to be any principal maxima of diffraction formed.

Since the values of m_x, m_y, and m_z are integral, and may hence change by only integral amounts, equation (7) will be satisfied by a number of discrete values of λ, and only by these values. When the incident beam is monochromatic, principal maxima of diffraction will be formed only if, by a rare chance, the wave-length happens to have one of this series of discrete values. Otherwise there will exist no direction in which a principal maximum can be formed—the waves must pass on in the direct beam.

When, as in the actual experiment of Laue, Friedrich, and Knipping, the

incident beam is heterogeneous, containing all wave-lengths within a somewhat indefinite range, there will be principal maxima of diffraction formed in those wave-lengths which satisfy equation (7), that is, in those wave-lengths for which three cones, one cone of each of the three families about each of the three co-ordinate axes, intersect in a line. Other wave-lengths pass on in the direct beam. Those wave-lengths for which three cones intersect in a line are diffracted along this line of intersection. Where the lines of intersection, for various discrete wave-lengths, meet the photographic plate, the so-called "Laue spots" are formed, which are illustrated in Figure 195*b*.

While theoretically m_x, m_y, and m_z may have any integral values, practically only small integral values are of importance. Maxima involving values of these integers higher than 5 or 6 are seldom observed.

It is to be noted that in the Laue experiment (in which there is a single direction of incidence) there is no "spectrum" formed, just as there is no spectrum formed in Bragg reflection when there is a single direction of incidence. That this should also be true in the Laue experiment is what we should expect in view of the circumstance, already explained in outline in section 6, that all Laue beams may be regarded as "reflected," according to Bragg's view, in one set or another of net planes. The fact that no spectrum is formed when there is a single direction of incidence is a fundamental characteristic of space-lattice diffraction.

18. Illustrative Examples.—The Laue theory requires three numbers, m_x, m_y, and m_z, to specify the order of diffraction. Each of these numbers represents the order of diffraction with reference to successive lattice particles lying along one of the axes of co-ordinates, that is, with reference to a specified direction in the crystal. The Bragg theory, on the other hand, requires only a single number to represent the order of diffraction, but requires in addition three Miller indices to specify the orientation of the set of crystal planes in which the reflection is taking place.

It will be instructive to select more or less arbitrarily several sets of values of m_x, m_y, and m_z, and study the diffraction which takes place in each case in detail.

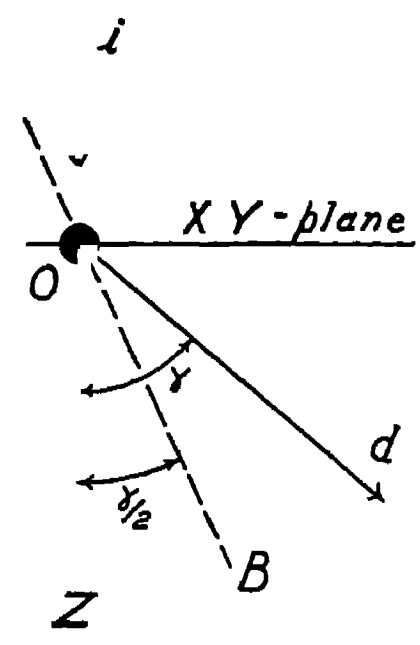


FIG. 217

In connection with the examples which are to follow it will be shown that a plane OB (Fig. 217), perpendicular to the plane of the figure, which passes through a lattice point O and bisects the angle of deviation γ of the diffracted beam, d, measured from the direction of incidence i, that this plane OB is one of a parallel set of net planes. These net planes may then be regarded in each case as reflecting planes. The plane OB will be referred to for brevity as the "bisecting plane."

For the order $m_x=3$, $m_y=0$, $m_z=-1$, the wave-length is determined by equation (7) to be $\lambda=a/5$, and by equations (4) we have $\cos\alpha=3/5$, $\cos\beta=0$ or $\beta=90°$, and $\cos\gamma=4/5$. These various numerical values are entered in the top row of Table V (the remainder of the table contains other

TABLE V

(LAUE)		(LAUE)			(BRAGG)	
Order m_x, m_y, m_z,	WAVE-LENGTH	Direction Cosines				
		$\cos\alpha$	$\cos\beta$	$\cos\gamma$	Order m	D
3, 0, −1	$\frac{1}{5}a$	$\frac{3}{5}$	0	$\frac{4}{5}$	1	$\frac{a}{\sqrt{10}}$
6, 0, −2 . .	$\frac{1}{10}a$	$\frac{3}{5}$	0	$\frac{4}{5}$	2	$\frac{a}{\sqrt{10}}$
9, 0, −3	$\frac{1}{15}a$	$\frac{3}{5}$	0	$\frac{4}{5}$	3	$\frac{a}{\sqrt{10}}$
2, 0, −1	$\frac{2}{5}a$	$\frac{4}{5}$	0	$\frac{3}{5}$	1	$\frac{a}{\sqrt{5}}$
4, 0, −2	$\frac{1}{5}a$	$\frac{4}{5}$	0	$\frac{3}{5}$	2	$\frac{a}{\sqrt{5}}$
1, 1, −1	$\frac{2}{3}a$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	1	$\frac{a}{\sqrt{3}}$

data to which we shall refer shortly). Figure 218 is drawn in the plane containing the incident and diffracted beams when the values of m_x, m_y, m_z

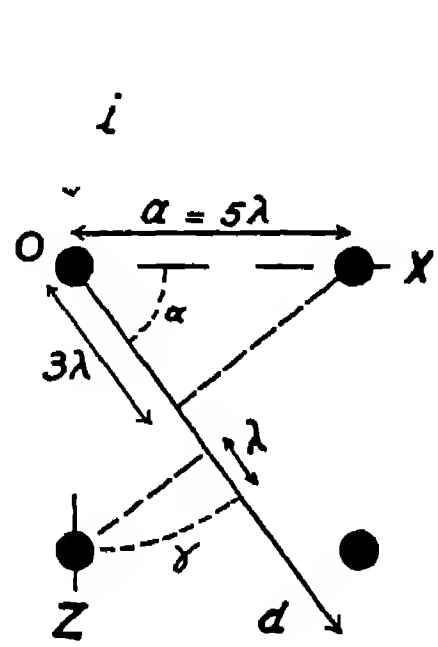


FIG. 218

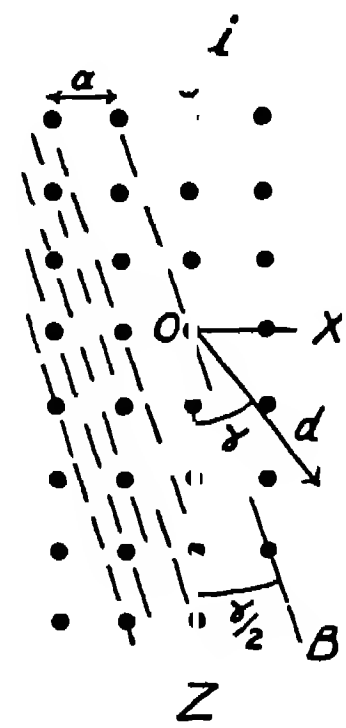


FIG. 219

are those just given; this plane now coincides with the XZ-plane. Path lengths λ and 3λ laid off along the arrow d bear an evident relation to order

of diffraction. Figure 219 illustrates the same case but is drawn to include a larger number of lattice points. The bisecting plane, OB, makes an angle $\gamma/2$ with the Z-axis. Invoking a well-known trigonometric formula, we have:

$$\sin\frac{\gamma}{2}=\pm\sqrt{\frac{1-\cos\gamma}{2}}=\sqrt{\frac{1-4/5}{2}}=\frac{1}{\sqrt{10}}. \tag{8}$$

When $\sin(\gamma/2)$ has the value $1/\sqrt{10}$, the line OB passes through the lattice point which is illustrated as lying on it. In fact, this line passes through a row of lattice points, and the plane OB, containing the Y-axis, is one of a set of net planes which have lattice points rectangularly arrayed upon them. These net planes, some of which are indicated in the figure by broken lines, may be regarded as reflecting planes. As will be shown, the minimal distance, D, between these net planes has the proper value to reflect the wave-length $\lambda=a/5$, according to Bragg's law, $m\lambda=2D\sin\theta$, when $m=1$. The angle θ is identical with $\gamma/2$, and thus $\sin\theta=\sin(\gamma/2)=1/\sqrt{10}$. Also, $D=(a/3)\cos(\gamma/2)$, as may be deduced from the figure. Since $\sin(\gamma/2)=1/\sqrt{10}$, we have $\cos(\gamma/2)=3/\sqrt{10}$, and hence $D=(a/3)(3/\sqrt{10})=a/\sqrt{10}$. Substituting these values of D and $\sin\theta$ in Bragg's law, we see that the law is satisfied and the diffracted beam may hence be regarded as reflected in the set of planes in question.

In the example given in the second line of Table V, the numbers giving the order $(6, 0, -2)$ are each twice as great as those giving the order in the first example; in the third example the numbers giving the order $(9, 0, -3)$ are each three times as great as in the first example. The wave-lengths are respectively one-half and one-third of the wave-length in the first example. The direction of diffraction is the same in all three cases and hence the same set of net planes act as "reflecting" planes in all three instances. The Bragg reflections are respectively of the second and third order. The common divisor of the numbers which give the order in the Laue theory is 2 in the second example and 3 in the third example. This common divisor is in each case the order m in Bragg's law.

The intercept ratios of the reflecting planes, in the present example, are evidently $1:\infty:-3$, and hence the Miller indices of the planes of this set are $(3, 0, -1)$. That is, the indices of the reflecting planes are the same as the numbers $m_x=3$, $m_y=0$, and $m_z=-1$, which give the order of diffraction in the Laue theory—for the greatest wave-length which is diffracted in the given direction. Or, stated differently, the Miller indices of the reflecting planes are equal to the numbers obtained by clearing the numbers giving the Laue order of diffraction of any common divisor which the latter may have. And the common divisor in question is equal to the order of Bragg reflection.

The various relationships arrived at in connection with the present example hold in all cases. These same relationships are deduced from a more general point of view in Appendix G.

In the example given in the fourth line of Table V the numbers giving the order are 2, 0, -1. Equation (7) now yields $\lambda = 2a/5$, and equations (4), (4′), and (4″) yield $\cos\alpha = 4/5$, $\cos\beta = 0$ or $\beta = 90°$, and $\cos\gamma = 3/5$, and it is evident from the geometry of the case that $\sin\theta = \sin(\gamma/2) = 1/\sqrt{5}$ and $D = a/\sqrt{5}$. The order of Bragg reflection is $m = 1$.

The Laue order 4, 0, -2 will be diffracted in the same direction as the order 2, 0, -1, just considered, but the wave-length which is diffracted will now be half as great, or will be $\lambda = a/5$, and the order of Bragg reflection will be $m = 2$.

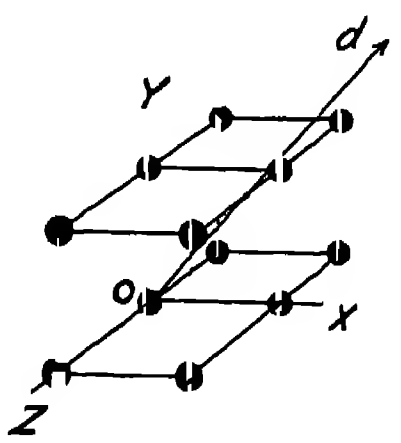


FIG. 220

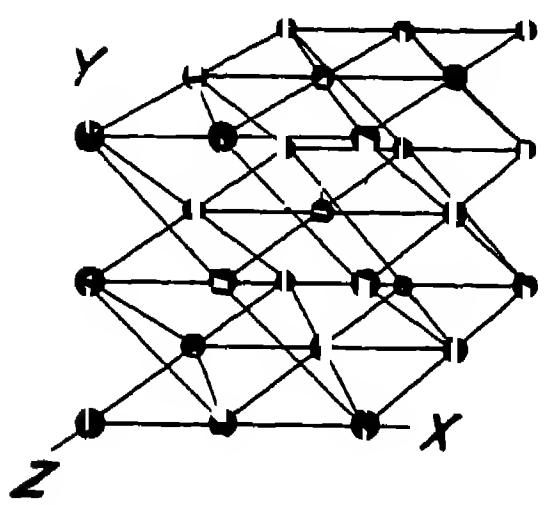


FIG. 221

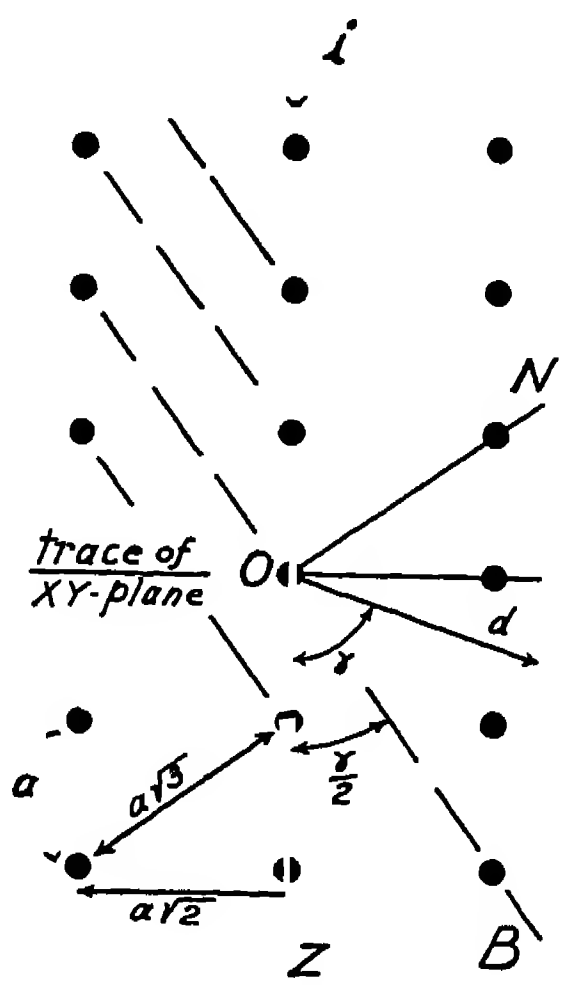


FIG. 222

In the sixth line of the table, the data for the order (1, 1, -1) are given, as deduced from equations (4), (4′), (4″) and (7). In Figure 220 the direction of diffraction is represented by the arrow d. The plane containing the incident and diffracted beams makes an angle of 45° with the X- and Y-axes. A diagram in this plane is drawn in Figure 222. The particles are separated by distances of a in the direction of the Z-axis and by $a\sqrt{2}$ in the direction perpendicular to this axis. For the bisecting plane:

$$\sin\frac{\gamma}{2} = \pm\sqrt{\frac{1-\cos\gamma}{2}} = \sqrt{\frac{1-\frac{1}{3}}{2}} = \frac{1}{\sqrt{3}}. \tag{9}$$

When sin $(\gamma/2)$ has this value, the line OB passes through the lattice points which are indicated as lying upon it. The normal to the bisecting plane, ON makes an angle with the Z-axis which will be denoted by ζ, and:

$$\cos\zeta = -\sin\frac{\gamma}{2} = -\frac{1}{\sqrt{3}}. \tag{10}$$

Hence ON passes through the lattice point having co-ordinates $(a, a, -a)$, and is a body diagonal of the unit cube. Referring to Figure 221, the diagonal set of net planes shown are normal to the body diagonal of the unit cube. One of the planes intersects the X, Y, and Z-axes respectively at distances from the origin equal to a, a, and $-a$. In comparing Figures 221 and 222 it must be remembered that Figure 222 is drawn in the 45° plane. The diagonal set of net planes evidently bisects the angle between the incident and diffracted beams, and these planes may be regarded as reflecting planes; the separation of the planes, $D = a/\sqrt{3}$, is the correct one to reflect the wave-length $\lambda = 2a/3$ in the order $m = 1$, according to Bragg's law, when $\sin\theta = \sin(\gamma/2) = 1/\sqrt{3}$, as it does in the present instance.

19. The Laue Diagram.—Let us suppose that Figure 223 occupies the plane of a photographic plate erected perpendicularly to the Z-axis, and that

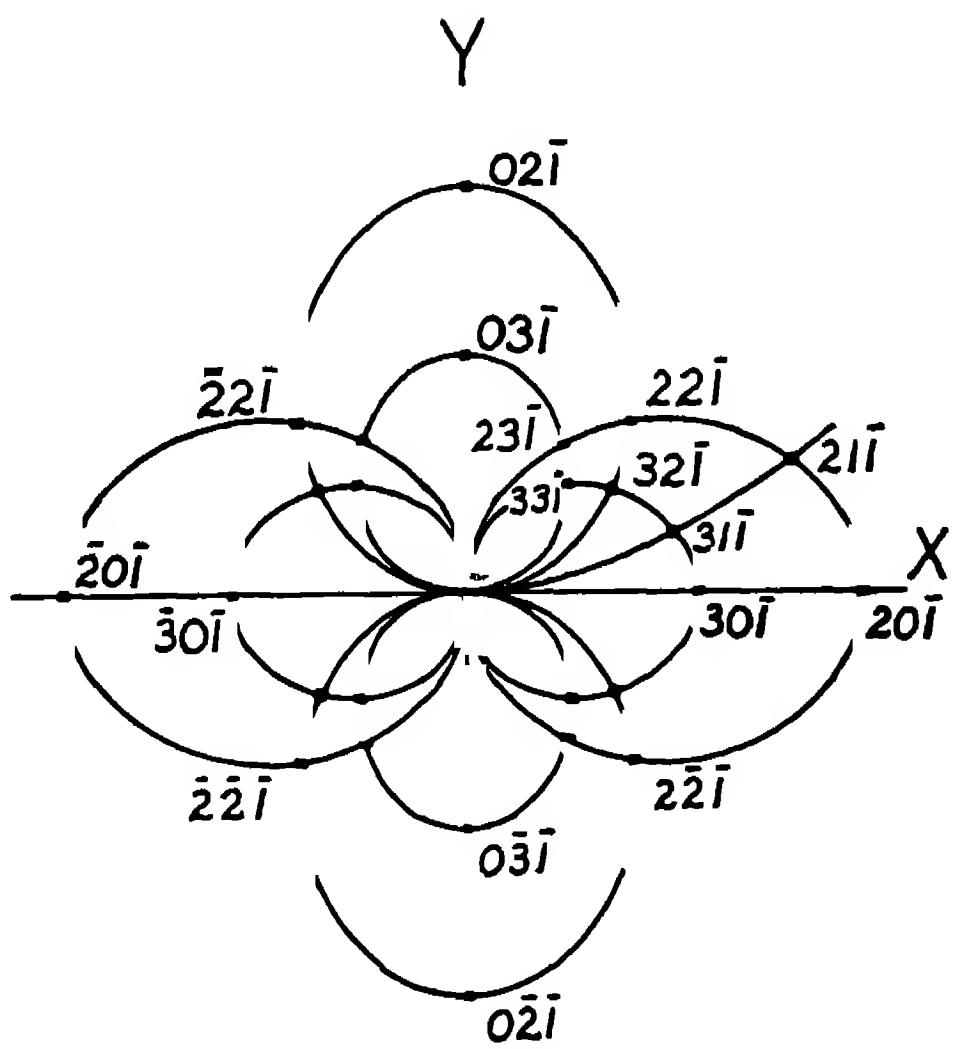


FIG. 223

the dot on the X-axis of the figure which is marked $30\bar{1}$ is the Laue spot due to diffraction of the order $m_x = 3$, $m_y = 0$, and $m_z = -1$, or briefly of the

order $30\bar{1}$, and that the dot marked $20\bar{1}$ is the Laue spot due to diffraction in the order $20\bar{1}$, etc. The numbers $30\bar{1}$, $20\bar{1}$, etc., with which the spots are designated give the Laue order of diffraction, and at the same time give the Miller indices of the planes in which the beam giving rise to the spot may be regarded as undergoing Bragg reflection. As deduced in the previous section and noted in Table V, the wave-length diffracted in the order $30\bar{1}$ is $a/5$, and the wave-length diffracted in the order $20\bar{1}$ is $2a/5$, or twice as great as that diffracted in the order $30\bar{1}$. In the scale of the figure the photographic plate would be at a distance of only 20 mm from the crystal.

The spots on the negative X-axis have the same order numbers as the corresponding ones on the positive side except for a reversal in the algebraic sign of the first order number. The order numbers of the spots on the Y-axis are deducible from those of the spots on the X-axis by interchanging the first and second numbers, as we should expect.

The spots occur at the various intersections of certain ellipses, of which a few are represented. The ellipses all pass through the origin of co-ordinates of the figure. Let us inquire into the reason why the spots occur on these ellipses.

Any row of particles of a lattice may be considered to determine a so-called "zone-axis," and all sets of net planes of the lattice which are parallel to this row of particles are said to belong to a common "zone" which has this row as its zone-axis. One plane of each of the sets belonging to the zone will pass through, or contain, the row of lattice particles.

Referring again to Figure 219, the row of particles lying along the line OB determines a zone-axis, and the net planes parallel to this line and perpendicular to the plane of the figure, of which some are indicated by broken lines, are one set of net planes of the zone determined by the direction OB. Other sets of net planes belonging to this same zone will be inclined to the plane of the figure but will have the same traces in the plane of the figure, and there will always be one plane of each set which contains the line OB. That is, if we imagine a geometric plane which passes through OB and rotates about OB as axis, this plane will become in turn a member of each of the sets of net planes of the zone determined by OB, and a few of these sets of planes will be sufficiently densely populated, and have sufficiently great separation, to have physical importance. These will have rather low Miller indices. Now whenever Bragg reflection occurs in any one of the sets of planes which belong to the zone, the reflected ray must make the same angle with the row of particles OB which the incident ray makes with it, namely, an angle $\gamma/2$, in the notation of the figure. Therefore, all of the diffracted beams must lie on a circular cone of angle $\gamma/2$ about the line OB, and this cone will intersect the photographic plate in an ellipse. The cone

will contain the Z-axis; hence the ellipse will pass through the point $x=0$, $y=0$, or, through the center of the pattern of spots.

The reader will recall that Figure 219 is drawn on the basis of the diffraction of the Laue order $30\bar{1}$. Or, in other words, the figure is drawn on the basis of Bragg reflection in the set of net planes having indices $(30\bar{1})$. The other sets of net planes which belong to the zone determined by the row of particles lying along OB also make intercepts along the X- and Z-axes which are in the ratio 1:3, and hence the physically important sets will each have indices of the form $(3k\bar{1})$, where k will take, in succession, for each set, one of the values 0, 1, 2, 3, etc. The other spots occurring on the ellipse of Figure 223 which passes through the spot $30\bar{1}$ are thus to be numbered $31\bar{1}$, $32\bar{1}$, $33\bar{1}$, etc., these numbers designating in each case, on the one hand, the indices of the planes in which Bragg reflection occurs, and, on the other, the Laue order of diffraction. The basic principles which underlie the arrangement of the spots in the Laue diagram represented in Figure 223 should now be clear.

The relative intensities of the spots are determined by the details of the structure of the crystal; for example, with a face-centered lattice some of the spots expected with a simple cubic lattice drop out altogether.

For further information regarding the interpretation of Laue photographs, see the references given below.*

We now change the subject of discussion rather radically.

20. Correction of Bragg's Law.—As stated in section 5, the velocity of X-rays in matter is greater than the velocity in vacuum by about 1 part in 10,000; the exact amount of difference is of course a function of the kind of matter and of the wave-length of the X-radiation. Because of the slight change in velocity which occurs as X-rays enter a crystal, Bragg's law, as we have hitherto applied it, is not rigorously correct. The correction which must be applied is very small, but since all possible accuracy is often required in scientific investigation, it becomes a matter of importance to inquire into the modification of Bragg's law which should be made on account of the minute change in the velocity of X-ray waves which occurs when these enter a crystal.

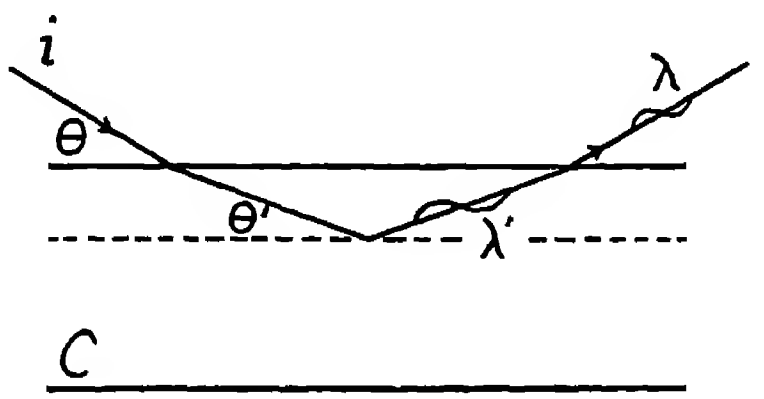


FIG. 224

Referring to Figure 224, let i be the incident beam and C a crystal. Upon entering the crystal, the beam is bent *away* from the normal. The broken

* W. H. and W. L. Bragg, *X-rays and Crystal Structure* (4th ed.; New York: Harcourt, Brace & Co., 1924); Wyckoff, *op. cit.*

horizontal line in C represents one of the reflecting net planes, which are supposedly parallel to the face of the crystal. The separation of these planes we shall, as previously, denote by D. It is evident that Bragg's law will hold rigorously *within* the crystal provided that we insert in this law the glancing angle which obtains within the crystal, which we shall denote by θ', and the wave-length which obtains within the crystal, which we shall denote by λ'. Thus:

$$m\lambda' = 2D \sin \theta'. \tag{11}$$

Let θ be the glancing angle and λ the wave-length outside of the crystal, that is, theoretically, in vacuum. The refraction which takes place at the surface of the crystal will follow the ordinary optical law of refraction, that is, Snell's law. Since we are now dealing with "glancing angles," that is, with the angles complementary to the angles of incidence and refraction, the law of refraction becomes:

$$\mu = \frac{\lambda}{\lambda'} = \frac{\cos \theta}{\cos \theta'}, \tag{12}$$

where μ is the index of refraction, which is slightly *less* than unity. By introducing equation (12) into equation (11), we obtain:

$$m\lambda = 2D\mu \sin \theta' = 2D[\mu^2 - \cos^2 \theta]^{\frac{1}{2}}. \tag{13}$$

Now let:

$$\mu = 1 - \delta, \tag{14}$$

where δ is the small amount by which the index of refraction is less than unity. Inserting this value of μ into equation (13) and remembering that when $(1-\delta)$ is squared the term in δ^2 may be neglected because it is small, we have:

$$m\lambda = 2D[\sin^2 \theta - 2\delta]^{\frac{1}{2}},$$

or, since δ is small:

$$m\lambda = 2\left[D\left(1 - \frac{\delta}{\sin^2 \theta}\right)\right] \sin \theta = 2D_m \sin \theta, \tag{15}$$

where by definition:

$$D_m = D\left(1 - \frac{\delta}{\sin^2 \theta}\right). \tag{16}$$

In equation (15) we have again arrived at an equation in terms of the wave-length and glancing angle outside of the crystal. Moreover, this equation has the same form as Bragg's law, but the constant D of Bragg's law is replaced by a quantity D_m which varies slightly with the glancing angle.

The wave-length which would be reflected at the glancing angle θ if Bragg's law held accurately is spoken of as the "Bragg wave-length" and is given by the equation:

$$m\lambda_B = 2D \sin \theta \,. \tag{17}$$

It is to be noted that the Bragg wave-length is *greater* than the true wave-length because D is greater than D_m.

Similarly, the glancing angle θ_B at which the true wave-length λ would be reflected if Bragg's law held is spoken of as the "Bragg angle," and this is given by the equation:

$$m\lambda = 2D \sin \theta_B \,. \tag{18}$$

And it is to be noted that the Bragg angle is *less* than the angle at which reflection actually occurs, because D is greater than D_m.

Since the correction which appears in equations (15) and (16), namely, $\delta/\sin^2 \theta$, is always quite small, an approximate value of $\sin \theta$ will suffice in evaluating it. Such a value is readily at hand in Bragg's law. The method of applying the correction will be presented later. Be it noted, meanwhile, that the correction $\delta/\sin^2 \theta$ can never be less than δ, because $\sin \theta$ can never be greater than unity.

Evaluations of δ have been made in three principal ways:

1. Direct experimental evaluations have been made by passing a narrow pencil of X-rays through prisms of glass, of quartz, of several metals, and of a few other materials. The small deviation of the beam which occurs may be recorded photographically, or may be measured on the double crystal spectrometer (this instrument will be discussed in sec. 24). In either case, prisms of large angle are used—usually of 90° or more—and the beam passes through very close to the vertex. The angle through which the beam is deviated by the prism is only some 2–6 seconds of arc (with λ from .7 to 1.5 Å). However, by causing either the incident or the refracted beam to graze the prism surface, the angle of deviation may be considerably increased, as in the optical analogue. But even then the angle of deviation becomes only a few minutes of arc. Values of δ ranging from 1×10^{-6} to 8×10^{-6} were found, in the foregoing wave-length range.

2. Experimental determinations of δ may also be made by an indirect method which will be outlined later.

3. Values of δ may be calculated from equations of dispersion theory.

The values of δ obtained in these three ways are all in substantial agreement.

According to dispersion theory:

$$\delta = \frac{e^2}{2\pi m} \sum \frac{N_i}{\nu^2 - \nu_i^2} \cdot \tag{19}$$

Here e and m are respectively the charge and mass of the electron, the charge being in electrostatic units. The significance of the summation term is as follows: ν is the frequency of the incident radiation in vibrations per second; thus $\nu = c/\lambda$, where c is the velocity of light and λ is the wave-length of the incident radiation. The subscript i is to take in succession the values 1, 2, 3, etc., and the summation is to be carried out with reference to i, in a manner to be further explained: Suppose the refracting material to be, for example, calcite ($CaCO_3$). When dispersion theory is applied in the form of equation (19), experiment indicates that one of the frequencies ν_i, call it ν_1, is to be given the frequency of the K absorption limit. There are two K electrons per atom of calcium, and from Avogadro's number and the molecular weight and density of calcite, we may calculate the number of the Ca K electrons per cubic centimeter of calcite. Let this number be N_1. In the same way let ν_2, ν_3, etc., be the natural frequencies of all other types of electrons present in calcite, and N_2, N_3, etc., the number of each type per cubic centimeter of the material. Taking our equation literally, the summation should now be extended over all types of electrons present, and this would be troublesome. But fortunately great simplification is possible in practice. Still considering calcite as the refracting material, the frequencies ν_i are, with one exception, very low compared to the frequency of the incident radiation when this is any but very soft X-radiation—which we shall rule out of discussion. The frequencies ν_i are with one exception so low that we may neglect ν_i^2 in comparison with ν^2 entirely. The exception is the frequency of the above-mentioned Ca K limit which, expressed in wave-length, falls at 3.064 Å.

When the incident radiation has a frequency which falls near any one of the natural frequencies of the refracting material, an effect of anomalous dispersion enters, and this effect should then, strictly speaking, be taken into account. However, the effect of anomalous dispersion caused by any one group of electrons is usually small, especially when the number of electrons in the group is small, as in the K group, which is, practically, the only one which ever need be considered. This effect may eventually be taken care of by a small correction when desired; in the meantime its existence may quite properly be overlooked entirely.

Be it noted parenthetically that equation (19) does not apply in the immediate vicinity of a natural frequency, because it does not take into account any effect of damping.

In view of the situation which has been outlined we may with moderately good approximation consider each and all of the frequencies ν_i practically equal to zero. Availing ourselves of the simplifications which thus become possible, we now proceed with the development of the theory:

Letting each $\nu_i = 0$ and taking the term ν^2, which is constant, outside of the summation, and, further, placing $\Sigma N_i = N$, where N thus becomes simply the total number of orbital electrons per cubic centimeter of the material,* we have:

$$\delta = \frac{e^2 N}{2\pi m \nu^2}. \tag{20}$$

Now substituting c/λ for ν:

$$\frac{\delta}{\lambda^2} = \frac{e^2 N}{2\pi m c^2}. \tag{21}$$

The right-hand side of this equation is constant, hence we may conclude that δ/λ^2 should be constant, or, rather, approximately constant; true constancy is not to be expected in view of the fact that the mathematical steps just taken became possible only upon neglecting the effect of the natural frequencies of the refracting material. The prediction that δ/λ^2 should be approximately constant is well borne out by experiment.

Since the entire correction to be applied to Bragg's law is small in any case, it becomes possible to develop a theory for applying the correction upon the assumption that δ/λ^2 is strictly constant, and then, subsequently, when desired, making a final very small correction for the lack of constancy of δ/λ^2 caused by the effect of natural frequencies of electron groups within the diffracting crystal. The final correction may be obtained from graphs which have been published, principally by W. Larsson. For reproductions of these and for a more comprehensive account of all that we are here presenting, as well as for detailed references, the reader is referred to M. Siegbahn's *Spektroskopie der Roentgenstrahlen* (2d ed.), chapter ii.

The assumption that δ/λ^2 is constant will apply especially well to crystals containing only light atoms, because for such crystals all of the natural frequencies are very low—falling almost without exception into the region of very soft X-radiation. Moreover, the group of crystals which con-

* Avogadro's number, 6.06×10^{23} per mol, times one-half of the molecular weight of the material gives the number of orbital electrons per mol, with good approximation. The number per cubic centimeter is then readily deduced.

tain only light atoms comprises practically all of the crystals ordinarily used for wave-length measurements. The situation is thus very fortunate.

We now proceed, again taking calcite as our specific example because it is of all crystals the most important in X-ray work.

The quantity δ/λ^2 is readily evaluated from equation (21) and proves to be equal to 3.69×10^{10} for calcite when λ is given in centimeters, and hence 3.69×10^{-6} when λ is in Ångstrom units.

Returning to the consideration of equations (15) and (16): As previously pointed out, an approximate value of $\sin\theta$ may be used in the correction term $\delta/\sin^2\theta$. Let us substitute the value of $\sin\theta$ given by Bragg's law, namely, $\sin\theta=m\lambda/2D$. Equation (15) then becomes:

$$m\lambda=2\left[D\left(1-\frac{\delta}{\lambda^2}\cdot\frac{4D^2}{m^2}\right)\right]\sin\theta=2D_m\sin\theta\,, \tag{22}$$

and introducing for δ/λ^2 the value given above, equation (16) becomes:

$$D_m=D\left(1-3.69\times10^{-6}\times\frac{4D^2}{m^2}\right). \tag{23}$$

From these equations it follows that the correction to be made in D to obtain D_m is the same for all wave-lengths, in the same order m, and that the correction is greatest in the first order, and decreases with the square of the order. Further, for high orders of short wave-lengths, the correction practically vanishes—because δ becomes very small for short wave-lengths; as pointed out in discussing equations (15) and (16), the correction term is never smaller than δ.

Denoting the values of D_m for $m=1$, 2, 3, etc., by D_1, D_2, D_3, etc., and remembering that the correction is one of subtraction, we have $D_1<D_2<D_3$, etc., and, as m approaches infinity D_m approaches D—because δ approaches zero as λ approaches zero—and it is only when λ approaches zero that m may approach ∞.

From equation (22) we have:

$$\frac{\sin\theta}{m}=\frac{\lambda}{2D_m}. \tag{24}$$

Consequently, when measurements are made upon the same spectrum line in different orders, the ratio $\sin\theta/m$ decreases slightly with increasing order —because D_m increases slightly with increasing order. In passing from order to order, beginning with $m=1$, the decrease in $\sin\theta/m$ is relatively large at first, and becomes rapidly smaller, approaching the value $\lambda/2D$ asymptoti-

cally for large values of m, when these are possible. But large values of m will of course be possible only when λ is small, and in this case δ is correspondingly small.

By making measurements upon the same spectrum line in different orders and plotting the resulting value of $(\sin\theta)/m$ as ordinate, against m, as abscissa, curves are obtained from which D_m may be evaluated for each order m, and from these values of D_m, one may then calculate δ for the given crystal and wave-length λ. This is the indirect experimental method spoken of above by which evaluations of δ may be made. The fact that values of δ determined by this method, for a beam which is undergoing diffraction, agree with values obtained by the prism method, or from dispersion formulae, for a beam not undergoing diffraction, is worthy of note. This agreement could not have been predicted offhand; but, it is called for by the dynamical theory of crystal diffraction.

C. G. Darwin* as early as 1914 arrived theoretically at the conclusion that Bragg's law requires correction and calculated the amount of the correction for the first-order reflection of Pt $L\beta$ radiation, $\lambda=1.11$ Å, from rocksalt. Not until 1919, however, was deviation from Bragg's discovered experimentally. In that year W. Stenström observed that the ratio $(\sin\theta)/m$ is not accurately constant, as it should be if Bragg's law held rigorously, but decreases with increasing order, as above described. The work of Stenström gave the first experimental evaluation of the quantity δ. Stenström's method was subsequently applied with greatly improved technique by A. Larsson, G. Hjalmar, and others.

The small deviation which an X-ray beam suffers in passing through a prism was not discovered until after Stenström's work, from which the discovery of the deviation caused by a prism followed as an indirect result.

The value of D for calcite is approximately 3.03 Å, and since $\delta/\lambda^2=3.69\times10^{-6}$, the value of $(\delta/\lambda^2)\cdot(4D^2/m^2)$ becomes $(135\times10^{-6})/m^2$. For calcite the law of reflection thus becomes:

$$m\lambda=2\left[D\left(1-\frac{135\times10^{-6}}{m^2}\right)\right]\sin\theta=2D_m\sin\theta\,. \tag{25}$$

It follows from this equation that the amount which must be subtracted from the Bragg wave-length to obtain the true wave-length is .135 X.U. when the wave-length is 1 Å, is .270 X.U. when the wave-length is 2 Å, etc. —when the measurements are in the first order. When in the second order, the amounts to be subtracted are 1/4 of the amounts given, when in the third order, 1/9, etc.

* *Phil. Mag.*, **27**, 315 and 675, 1914.

For each given kind of crystal, and reflection from a given face, D_m is the same for all wave-lengths in the same order, hence the values of D_1, D_2, D_3, etc., are, for convenience, tabulated for each of the more important crystals used in wave-length measurements, reflecting from the face commonly used. By way of example, such a tabulation is presented in Table VI

TABLE VI

FOR REFLECTION FROM A (100) FACE OF CALCITE, AT 18° C.

$D_1 = 3029.04$ X.U.
$D_2 = 3029.34$
$D_3 = 3029.40$
$D_4 = 3029.42$
$D_5 = 3029.43$
. . . . . .
As $m \doteq \infty$, $\lim D_m = D = 3029.45$

for calcite at a temperature of 18°, reflection being supposedly from a (100) face. The manner in which the precise values given in this table were arrived at is roundabout and requires explanation:

Long before there were any experimental data whatever upon which to base corrections to Bragg's law, W. H. Bragg calculated the value of D for rocksalt from the best value of Avogadro's number then available and the measured density of rocksalt. The value at which he arrived for reflection from a (100) face was $D = 2814$ X.U., the last figure being uncertain by one unit or more. Soon, however, relative measurements of X-ray wave-lengths could be made with an accuracy far exceeding the accuracy with which D was known for any crystal. Accordingly, since it was imperative to have a definite basis for comparing wave-lengths as accurately as possible, the value of D for rocksalt was taken as 2814.00 X.U., the two zeros following the decimal point being arbitrarily added as a matter of common agreement in order that the necessary, accurate basis for comparison of wave-lengths might be at hand. Most of the measurements were made in the first order. Hence we should really say today that it was long ago agreed by convention that for rocksalt the value 2814.00 X.U. should be used for D_1, and not for D; the supposedly true value, D, in this partially conventional scale would be somewhat higher. Furthermore, careful measurements made by numerous investigators upon some of the same spectrum lines, also in the first order, led to the conclusion that for calcite what we now call D_1 should be taken as 3029.04 X.U. in order to maintain the same scale of wave-lengths arrived at on the basis of taking D_1 for rocksalt as 2814.00 X.U. As experimental technique improved, it developed that calcite was greatly superior as a crystal to rocksalt, but in the absence of knowledge of the lattice constants of calcite or of

any other crystal with the absolute accuracy which was required, namely, .0003 per cent, the above value of 3029.04 X.U. of D_1 for calcite was adhered to. From this the other values given in Table VI were then derived.

Using the values given in Table VI as a basis, especially $D_1 = 3029.04$, because most of the measurements were in the first order, the wave-lengths of three important X-ray lines were determined with the utmost care in the year 1919 by M. Siegbahn, who found the values given in Table VII.

TABLE VII

Fe	$K\alpha_1$	1932.076 X.U.
Cu	$K\alpha_1$	1537.396
Mo	$K\alpha_1$	707.831

These constitute the standards upon which X-ray wave-lengths are now based. A host of wave-length measurements have been made which are based upon them. They are what would be true wave-lengths, in vacuum, that is, corrected for the error of Bragg's law, if the value of D for calcite at 18° C. were accurately 3029.45 X.U., as given in Table VI. However, one must remember that the value of D is not known absolutely for calcite or for any other crystal to nearly the degree of accuracy which the number of significant figures which are carried implies. The value may well be in error by 1 X.U., or .03 per cent, and quite possibly by a greater amount.

In the year 1925, Compton, Beets, and De Foe, using a supposedly more accurate value of Avogadro's number than Bragg had available a decade earlier and also a more accurate value of the density of the crystal, calculated the value of D for calcite at 20° C. as 3029.1 ± 1 X.U. This reduces, at 18° C., to $D = 3029.04$ X.U., and $D_1 = 3028.63$ X.U., one further significant figure being arbitrarily carried. Accordingly, when, for some especial reason, wave-lengths are to be based upon the value of D given by Compton, Beets and De Foe, these will be lower by .0132 per cent than wave-lengths based upon the standards of Siegbahn, given in Table VII.

Instead of correcting for the error in Bragg's law by using a corrected value of D, a correction may be applied to the glancing angle θ. A small angle $\Delta\theta$ must then be subtracted from each true glancing angle θ, but this procedure is on the whole less convenient and hence less commonly resorted to than the method above outlined. However, details of carrying it out will appear, incidentally, in the course of developing the subject of the next section.

The theory of correcting for error in Bragg's law which we have considered applies in the case which is of greatest general importance, namely, when the face of the crystal is parallel to the planes from which the reflection is taking place—to the case of symmetrical reflection—and only to this case.

By expressly preparing a face at such an angle with respect to the reflecting planes that either the incident or the emergent beam will graze the face of the crystal, the effect of refraction may be considerably increased—and this method has upon occasion been made the basis of investigations upon the refraction of X-rays.

Be it also pointed out that when an X-ray beam is reflected as in the spectrograph of Rutherford and Andrade, from planes perpendicular to the faces of a crystal having parallel sides, no correction to Bragg's law is required. A beam entering a crystal is, under any circumstances of X-ray refraction, deviated away from the normal to the refracting surface. Now, in the case under consideration, the glancing angle made with the reflecting planes is increased by the refraction—and at the same time the wave-length is increased upon entering the crystal. It may be readily shown that the two effects compensate for each other and that Bragg's law may hence be used without correction, even for accurate wave-length measurements, when the reflection is from planes perpendicular to the faces of a crystal having parallel sides.

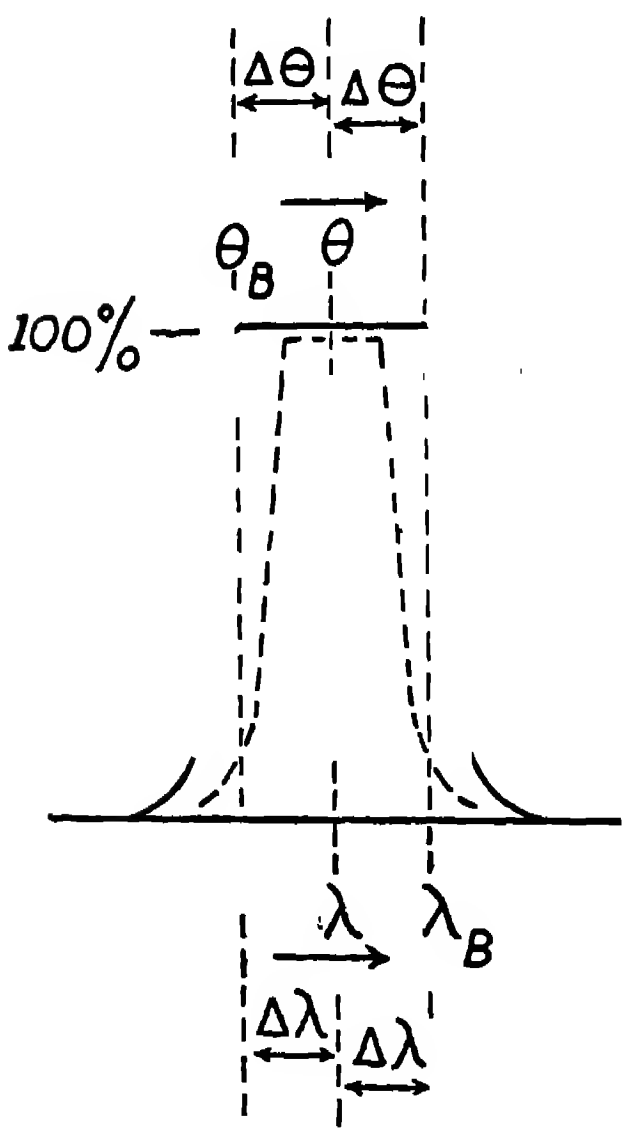


FIG. 225

21. Angle of Tolerance.—It was mentioned in section 5 that when an X-ray beam is reflected by a given set of crystal planes, reflection does not occur only at a unique angle θ, but rather over a small finite range of angle. In the dynamical theory of C. G. Darwin* and of P. P. Ewald,† a certain small range of angle $\Delta\theta$ is determined (see Fig. 225). For the kind of crystal which is postulated, the reflection would be *total* over a range extending from an angle $\Delta\theta$ below the unique value of θ which is given by the law of reflection in its corrected form, to an angle $\Delta\theta$ above θ. But the kind of crystal which is postulated in arriving at the value of $\Delta\theta$ (to be given presently) is really a decidedly non-physical mathematical abstraction. The crystal is not only assumed to be perfect in the sense of having no mosaic structure and no warping of the crystal planes, but all of the diffracting material is assumed to lie in successive layers of zero width. The effect of structure factor is thus neglected. In actual crystals there is a continuous periodic distribution of the diffract-

* *Ibid.*

† *Phys. Zeit.* **26**, 29, 1925.

ing material in a direction normal to the crystal planes, the density of this material not falling to zero anywhere between these planes. The crystal is also assumed to be non-absorbing. The magnitude of the angle $\Delta\theta$, for the postulated crystal, is shown to be:

$$\Delta\theta = \frac{\delta}{\sin\theta\cos\theta} = \frac{2\delta}{\sin 2\theta}. \tag{26}$$

The solid curve of Figure 225 represents the intensity of reflection, which has a constant value of 100 per cent over the range $2\Delta\theta$ and falls rapidly to zero outside of this range.

The modification which is introduced by having the diffracting material distributed as it actually is distributed, and the modification which is introduced by absorption of the radiation within the crystal will be spoken of later.

The angle over which a given wave-length λ is reflected by a crystal is called the "angle of tolerance" of the crystal for that wave-length. Accordingly, we may say that $2\Delta\theta$ is the theoretical angle of tolerance for the postulated non-physical crystal—or $\Delta\theta$ is the half-angle of tolerance.

Now it may be readily shown that the small angle $\Delta\theta$ equals the angle $\theta-\theta_B$, that is, the difference between the angle at which reflection of the wave-length λ occurs according to the corrected law of reflection, for example, equation (15), and the Bragg angle θ_B calculated from Bragg's law, equation (18). This result may be deduced by subtracting equation (18) from equation (15), in which case one obtains:

$$\sin\theta - \sin\theta_B = \frac{\delta}{\sin\theta}. \tag{27}$$

Applying the formula for the difference of two sines, and remembering that since θ and θ_B differ but little from each other, one may write by way of approximation $\cos[(\theta+\theta_B)/2]=\cos\theta$ and $\sin[(\theta-\theta_B)/2]=(\theta-\theta_B)/2=\Delta\theta/2$. The value $\delta/\sin\theta\cos\theta$ is now readily obtained for $\theta-\theta_B=\Delta\theta$, in accord with equation (26). Thus the "Bragg angle" lies as indicated in Figure 225 at the lower limit of the range $2\Delta\theta$ over which total reflection supposedly occurs, the true angle θ occupying the central position of this range.

Further, by introducing the value of $\Delta\theta$ given in equation (26) into the law of reflection differentiated, $m\Delta\lambda=2D_m\cos\theta\cdot\Delta\theta$, one obtains an increment of wave-length $\Delta\lambda$ (see Fig. 225) given by the equation:

$$\Delta\lambda = \frac{2\delta D_m}{m\sin\theta} = \frac{2\delta D}{m\sin\theta} \text{ approximately.} \tag{28}$$

That is, all wave-lengths lying within a small range extending from $\lambda - \Delta\lambda$ to $\lambda + \Delta\lambda$ would be reflected by the postulated non-physical crystal when the glancing angle has the discrete value θ.

By subtracting equation (15) from equation (17), one may now obtain for the difference between the Bragg wave-length and the true wave-length $\lambda_B - \lambda = \Delta\lambda$, the value given in equation (28). That is, the Bragg wave-length λ_B lies as indicated in Figure 225 at the upper limit of the small range, $2\Delta\lambda$, of the wave-lengths which would be totally reflected at the glancing angle θ by the postulated crystal.

In order to convey a definite idea of the magnitudes of the quantities $\Delta\theta$ and $\Delta\lambda$, these quantities are plotted in Figure 226 as ordinate, respectively in seconds of arc and in X.U., against values of sin θ and of λ in Ångstroms, plotted as abscissa. The values given are what they would be for first-order reflection from a (100) face of calcite. The assumption formerly made that δ/λ^2 is constant and equals 3.69×10^{-6} is again introduced in plotting these curves. Be it repeated that $\Delta\theta$ represents, on the one hand, the half-value of what the angle of tolerance would be if all of the diffracting material were concentrated in the crystal planes, and represents, on the other hand, the amount by which the actual glancing angle exceeds the Bragg angle. And $\Delta\lambda$ represents, on the one hand, the half-value of the small range of wave-lengths which would be reflected by the postulated crystal at a given glancing angle, and represents, on the other hand, the amount by which the Bragg wave-length exceeds the true wave-length.

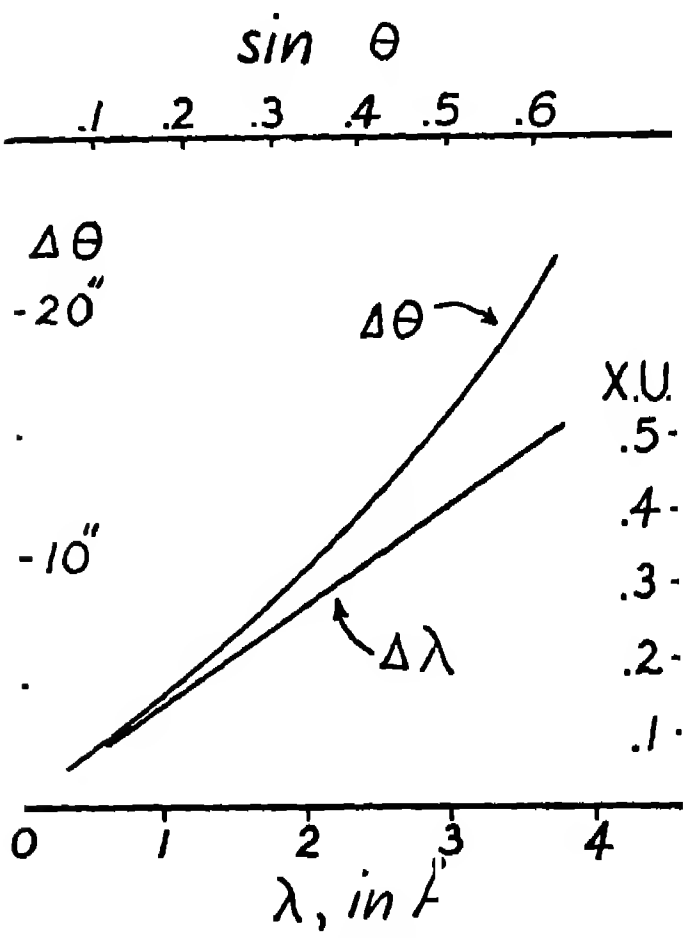


FIG. 226

As for the effect of structure factor—that is, the effect introduced by the continuous periodic distribution of the diffracting material between the crystal planes in place of the hypothecated concentration in layers of zero width—this effect is to cause the region of total reflection to contract in width. In the case of the foregoing supposed reflection from calcite, the region of total reflection reduces to about half of the width previously considered (see dotted curve of Fig. 225). Thus, the Bragg angle, θ_B, and the Bragg wave-length, λ_B, fall, as illustrated, well outside of the region of reflection for an actual crystal which is yet supposedly perfect in the sense of having no mosaic structure or warping of the crystal planes. However, mo-

saic structure or warping may widen the range of angle over which reflection occurs very decidedly.

Finally, when absorption is taken into account, following J. A. Prins,* the curve obtained falls within the same limits as it does in the absence of absorption but is lower and is rounded off at the upper corners. Depending upon what assumptions are made in the theory, the curve may also be rather unsymmetrical between the limits within which reflection occurs.

22. The Diffraction of X-rays by Ruled Gratings.—First, a few words concerning the related subject of the specular reflection of X-rays by a polished surface. As stated in section 1, Roentgen found that X-rays are diffusely reflected, or scattered, from within the body of any material, but he failed to find specular reflection from any surface. Others who sought for specular reflection also failed to find it, until after nearly three decades the condition under which specular reflection occurs was discovered by A. H. Compton.†

Compton was led to his method of attacking the problem by the results of Stenström, which, as we learned in section 20, revealed that in the Bragg reflection of a given spectrum line by a crystal, the ratio $(\sin\theta)/m$ decreases slightly with increasing spectral order, as is to be expected if matter has an index of refraction for X-rays which is slightly less than unity. Accordingly, Compton reasoned, while X-rays are not specularly reflected by any surface when the beam is directed at the surface at an ordinary angle, for example, 45°, if X-rays should be directed at a polished surface at nearly grazing incidence, a "totally reflected" beam may be formed. Compton, using first heterogeneous radiation, and then homogeneous L-radiation from a tungsten target, $\lambda = 1.28$ Å, collimated the radiation by a succession of slits and directed it toward a plane surface of plate glass at very small glancing angles. Employing an ionization chamber, he searched for the specularly reflected beam and found it, and was able to prove that its integrated intensity was nearly equal to that of the incident beam so long as the glancing angle of incidence was somewhat less than 10′ of arc. At the glancing angle of 10′ the intensity had a steep downward gradient and at higher angles fell rapidly to zero. He thus determined the critical glancing angle for the glass which he was using to be about 10′, and this indicated an index of refraction less than unity by about $\delta = 4.2 \times 10^{-6}$, for $\lambda = 1.28$ Å. Corresponding experiments with a silver reflecting surface yielded 22.5′ for the critical angle and an index of refraction less than unity by about $\delta = 21.5 \times 10^{-6}$. These results accorded with calculations for the corresponding indices of refraction deduced from the theoretical dispersion formula given in equations (20) and (21).

* *Zeit. f. Phys.*, **63**, 477, 1930.

† *Phil. Mag.*, **45**, 1121, 1923.

The above-described research established the fact that specular reflection of X-rays occurs at very small glancing angles, and only at such angles. Following up this result, Compton and Doan,* in the year 1925, succeeded in diffracting X-rays by a ruled grating. They used a plane grating lightly ruled upon speculum having a comparatively large grating space, 1/500 cm, in order that several diffracted spectral orders might appear within the critical angle. They allowed radiations from successively a copper and a molybdenum target, after collimation by a succession of slits, to be incident upon the grating at glancing angles of some 10′ of arc, and erected a photographic plate roughly perpendicular to the expected emergent beams. These beams, making only small angles with each other, would all fall upon the plate simultaneously.

Their plates showed the central image strongly developed and, in addition, and weaker, the first so-called "inside" order and the first, second, and third so-called "outside" orders. The term "inside orders" is used to designate those diffracted images which lie between the central image beam and the grating surface, and the term "outside orders" is used to designate those which lie on the other side of the central image beam. The inside orders are usually designated by negative numerals and the outside orders by positive ones. The fact that only the first inside order was found by Compton and Doan is readily explained. The second and higher inside orders will be formed only when the path difference by way of successive lines of the grating may attain the values 2λ, 3λ, etc. as the direction of diffraction sweeps out the angle between the central image beam and the direction of grazing emergence. Under the conditions selected by Compton and Doan for their experiments, the path difference attained values as high as λ and higher, but not as high as 2λ. Hence the second and higher inside orders could not appear. Such spectral orders may, however, be recorded when larger glancing angles of incidence may be used, or when sufficiently coarse gratings are used.

In making calculations of X-ray wave-lengths diffracted by a grating it is convenient to express the grating law in a form differing from the usual one. The usual form is $m\lambda = a\ (\sin\theta_1 + \sin\theta_2)$, where θ_1 and θ_2 are respectively the angles of incidence and of diffraction—measured from the normal to the grating surface. This may be written:

$$m\lambda = \pm a\ (\cos\varphi_1 - \cos\varphi_2)\ , \tag{29}$$

where φ_1 and φ_2 are the corresponding glancing angles. Let α be the angle between the diffracted beam and the central image beam (taken positive

* *Proc. Nat. Acad.*, **11**, 598, 1925.

when $\varphi_2 > \varphi_1$). Then $\varphi_2 = \varphi_1 + \alpha$. Since φ_1 is a small angle it is permissible to expand its cosine in series and retain only the first two terms, writing $\cos \varphi_1 = 1 - \varphi_1^2/2$; and $\cos(\varphi_1 + \alpha)$ may be treated in the same way. Then the grating law becomes, to sufficient approximation:

$$m\lambda = a\left(\varphi_1\alpha + \frac{\alpha^2}{2}\right). \tag{30}$$

Other modified forms of the grating law have also been used.

It should perhaps be mentioned that in the year previous to Compton and Doan's success, N. Carrara had attempted to diffract X-rays by a ruled grating, using small glancing angles, but he found only the specularly reflected beam.*

For making wave-length determinations, Compton and Doan desired to use monochromatic radiation. They secured this, using a molybdenum target, by diffracting the radiation first from a crystal at the proper angle to give the $K\alpha_1$ line, $\lambda = .708$ Å, and introduced this radiation into the grating spectrograph. The wave-length for this line, which they then calculated from the grating space of their grating and the measured angles of incidence and diffraction, agreed to within their errors of measurement, about .5 per cent, with the wave-length as determined from crystal measurements. But the work of Compton and Doan, which was undertaken as a search for new possibilities of diffracting X-rays, did not aim at the highest possible accuracy.

J. Thibaud shortly followed Compton and Doan in experimenting upon the diffraction of X-rays by ruled gratings and has continued making significant contributions in this field from time to time.†

A few words may to advantage be added at this point regarding the principles which underlie the specular reflection of X-rays:

Since the index of refraction of matter for X-rays differs so little as it does from unity, the intensity of any specularly reflected beam which might be formed at ordinary angles of incidence would perforce be extremely feeble, even if such a beam were formed. Now the question of the smoothness of the reflecting surface also enters in determining whether or not a specular beam will be formed. Even a highly polished surface is rough for wave-lengths as short as those of X-rays, which are of only the order of magnitude of atomic dimensions. At large glancing angles this roughness will be very damaging, but at small glancing angles it will introduce only relatively small path differences. This is in accord with the well-understood optical

* *Il nuovo cimento*, **1**, 107, 1924.

† For summarizing accounts see J. Thibaud, *JOSA and RSI*, **17**, 145, 1928; *Phys. Zeit.*, **29**, 241, 1928.

phenomenon that a surface too rough to exhibit specular reflection at ordinary angles may yet reflect specularly at nearly grazing angles. Hence X-rays will be specularly reflected at small glancing angles by a highly polished surface, whereas such a surface is too rough to reflect X-rays at ordinary angles.

If the reflecting medium (of index less than unity) were perfectly transparent, the reflection would remain total from grazing incidence up to the critical angle and then fall off rapidly. But since, actually, the second medium absorbs the radiation, the intensity of the reflected beam falls off gradually before one reaches what would be the critical angle in the absence of absorption. There is thus, strictly speaking, no critical angle. However, with hard X-rays this gradual falling-off is much less marked, and the critical angle is much better defined, than with soft X-rays. But for soft X-rays, on the other hand, the index of refraction differs much more greatly from unity than for hard X-rays, and consequently greater glancing angles of incidence may be used for soft X-rays. With glass, for example, at $\lambda = 45$ Å, a glancing angle of incidence as high as 3° may be used, and for $\lambda =$ 65 Å a glancing angle of 4.5° may be used. These limiting angles are, however, only about one-half of what the critical angles would be in the absence of absorption. The question, closely related to the one we have been discussing, of how and when the diffracted images from a grating disappear, as a function of increasing glancing angle, wave-length, spectral order, and positive or negative position of the diffracted image, has been studied to some extent by J. Thibaud* and by J. A. Prins.† Pertinent data have also been contributed by E. Dershem.‡

In ruling gratings for the diffraction of X-rays it is impossible to follow the principle applied for ordinary optical gratings, namely, of ruling the groove face at that angle which will cause a maximum of energy from each groove face to be reflected into a desired direction. By way of alternative, gratings for X-rays are usually ruled so lightly that a portion of the original surface remains between successive grooves. Under these circumstances we may, in approximation with the truth, consider the grooves as contributing no energy at all to the central and diffracted images, and, accordingly, consider the intervening unruled portion as contributing all of the energy. We are then confronted with the case of a grating consisting of alternate opaque and transparent strips except that the granting now acts by reflection. The most favorable width of groove is, on this hypothesis, such that the ruled and unruled portions are of equal width (see chap. 6, sec. 15). It is difficult,

* *Helvet. Phys. Act.*, **2**, 271, 1929. † *Nature*, **124**, 370, 1929.

‡ *Phys. Rev.*, **34**, 1015, 1929; also Dershem and Schein, *ibid.*, **37**, 1246, 1931, and *Zeit. f. Phys.*, **75**, 395, 1932.

however, to determine, from microphotographs, how nearly this supposed optimum condition obtains in the actual gratings used for X-ray work.

M. Siegbahn* has developed a ruling machine for the especial purpose of ruling very small gratings for X-rays. This machine rules gratings up to only 3 mm wide and with a length of ruled line of only 15 mm, but it is capable of ruling gratings up to 1800 lines per millimeter (about 45,000 lines per inch), which still perform excellently. The grooves ruled by Siegbahn's machine are moreover much "cleaner" than those ruled on the machines constructed for ruling larger, optical gratings.

For X-ray work, gratings on glass are preferred to those on speculum because they give greater intensity.

The diffraction of X-rays by the ruled grating has been resorted to in numerous researches, and, generally speaking, for either one of two purposes:

One of these purposes has been to obtain accurate absolute and independent determinations of certain X-ray wave-lengths, and compare these with the best determinations made upon the crystal spectrometer. And it has been found that wave-lengths determined respectively by gratings and by crystals differ by a larger amount than can be accounted for. Numerous researches have been undertaken in the attempt to clear up the cause of the discrepancy. Wave-lengths measured when using crystals are calculated from the measured angles of Bragg reflection and the lattice constant of the crystal—then a correction amounting to about .01 per cent in the first-order spectrum is made for deviation from Bragg's law. The absolute accuracy of these determinations is limited largely by the accuracy with which the lattice constant is known. This constant, for a given crystal, is calculated from the mechanical density of the crystal and Avogadro's number, the crystal form, of course, being taken into account. The quantities in question can each be determined, supposedly, within a few hundredths of 1 per cent. And the grating determinations can be made supposedly to within less than .1 per cent. Yet the two sets of independent determinations differ by from .1 to .3 per cent, the grating values being always the larger. Since Avogadro's number is an important constant and the best method of determining it involves other important constants, especially the electronic charge, numerous questions are raised—all by the above-mentioned discrepancy: May unknown errors have crept into the determination of the electronic charge? Or may the effective lattice constant of a crystal differ from that obtained by calculation, perhaps because of mosaic structure? Or may conditions obtaining at the surface of a crystal be different from those in the interior? Or may the grating law not hold accurately under the

* See Siegbahn and Magnusson, *Zeit. f. Phys.*, **62**, 435, 1930.

rather extreme conditions which apply when X-rays are diffracted by a ruled grating? Or may systematic errors have crept into the grating measurements?

The first investigator who made grating measurements of sufficient accuracy to disclose that grating determinations of wave-lengths are slightly but definitely higher than crystal determinations was E. Bäcklin, in a Doctor's thesis. Subsequently the most extensive and careful determinations have been made by J. M. Cork,* and by J. A. Bearden.†

The other principal objective, in using gratings, has been to investigate the spectra of various elements in the very soft X-ray region, 20–140 Å, where ordinary stable crystals cannot be used because they have insufficiently large lattice constants, and organic crystals which have large-enough lattice constants are unsatisfactory principally because of their lack of permanence.

The pioneer in the use of the plane grating in the soft X-ray region was J. Thibaud, and in the use of the concave grating was T. H. Osgood. Subsequently, important contributions have been made especially from Siegbahn's laboratory.‡

23. The Diffraction of X-rays by a Slit.—Following the announcement of Roentgen's discovery, in the year 1895, several investigators promptly attempted to diffract X-rays, especially by means of a slit. The first results were, however, all either negative, indicating that X-rays are propagated, as nearly as could be determined, rectilinearly, or when positive effects were at first believed to have been obtained, they eventually proved to have been introduced by the apparatus.§

Later and more thoroughgoing experimenters fared no better within the next fifteen years. Haga and Wind,‖ during the years 1899–1903, performed a series of careful experiments in which radiation from an X-ray tube passed first through a slit of .015 mm width and 4 mm length which served as a linear source, and then, at a distance of 75 cm, through the diffracting slit which tapered from a width of .025 mm at the broad end to zero at the other end. The length of this slit was 18 mm. A photographic plate was erected at a distance of 75 cm beyond the diffracting slit. While no pattern showing diffraction fringes or other fine structure was found, the investiga-

* *Phys. Rev.*, **35**, 1456, 1930; also Witmer and Cork, *ibid.*, **36**, 743, 1932.

† *Proc. Nat. Acad.*, **15**, 528, 1929; *Phys. Rev.*, **37**, 1210, 1931.

‡ For further information see J. Thibaud, *Phys. Zeit.*, **29**, 241, 1928; T. H. Osgood, *Phys. Rev. Suppl.*, **1**, 228, 1929; and Siegbahn, *op. cit.*, pp. 393 ff. Osgood's concave grating mounting is briefly described in our chap. 6, sec. 33.

§ See, e.g., *Comptes rend.*, **122**, 1896; J. Perrin, p. 187; G. Sagnac, p. 783; and Calmette and Lhuilier, p. 877; also L. Fomm, *Ann. d. Phys.*, **59**, 350, 1896.

‖ *Wied. Ann.*, **68**, 884, 1899; *Ann. d. Phys.*, **10**, 305, 1903, etc.

tors believed to have proved that the blackening of the photographic plate was broader relatively to the width of the geometrical shadow, at the narrow end of the slit, than at the broad end, as would be expected if diffraction had taken place. But B. Walter* questioned the validity of the conclusions of Haga and Wind, claiming that the relative broadening which they had measured was of photographic origin. Walter had also performed careful experiments of his own. In 1908 and 1909 Walter and Pohl,† still claiming that the broadening observed by Haga and Wind was spurious, repeated the experiments with even greater care, and concluded that there was no measurable real effect. The question was one of importance at the time, for the possibility of diffracting X-rays in any manner had not as yet been established.

In the year 1912 P. P. Koch‡ made careful photometric measurements upon the plates of Walter and Pohl, determining the rate at which the blackening fell off in passing from the center of the image to the side, and his measurements finally revealed that the gradient was slightly greater when the X-ray tube had been very "hard" than when it had been relatively "soft." The measurements of Koch were used by A. Sommerfeld§ to estimate the wave-length, or what was at the time called the "breadth of the pulse" of the radiation. Sommerfeld's estimate was .4 Å, which we now know must have been about the effective wave-length of the continuous radiation emitted by the rather "hard" X-ray tube which was used in the experiments.

All investigators had used heterogeneous radiation, because no method of isolating homogeneous radiation was then known.

The reader will recall that Laue's discovery of the diffraction of X-rays by crystals was made in the year 1912. Thenceforth the question of whether it might be possible to diffract X-rays by a slit lost some of its former importance.

In the year 1924 B. Walter|| made the first attempt to diffract monochromatic, or approximately monochromatic, radiation by means of a slit. He used the same apparatus which he had used years earlier in his experiments carried out in collaboration with Pohl, and used Cu K radiation of wave-length about 1.5 Å. This time he obtained actual diffraction fringes—only two or three faint ones—and these were of course very close together; but he was able to show that their spacing was that to be expected for the wave-length used. I. I. Rabinov¶ reported similar results in the following year.

* *Phys. Zeit.*, **3**, 137, 1902.

† *Ann. d. Phys.*, **25**, 715, 1908; **29**, 331, 1909.

‡ *Ibid.*, **38**, 507, 1912.

§ *Ibid.*, p. 473.

|| *Ibid.*, **74**, 661; **75**, 189, 1924.

¶ *Proc. Nat. Acad.*, **11**, 222, 1925.

Much finer photographs showing many distinct fringes were obtained in the years 1928–30 in Siegbahn's laboratory, in succession by E. Bäcklin, A. Larsson, and G. Kellström, using Al K radiation, $\lambda = 8.3$ Å and also other wave-lengths up to 45 Å. With these radiations the fringes are of course more readily obtained than with the much shorter wave-length of 1.5 Å used by Walter. However, even so, a magnification of some thirty diameters is required to render the photographed fringes readily distinguishable. Some of the photographs obtained in Siegbahn's laboratory were by-products of investigations carried out for a different primary purpose. In taking the photographs the diffracting slit was of uniform width, around .01 mm, and the distance from the slit to the photographic plate was around 50 cm. Numerous reproductions are given by Siegbahn.*

Be it mentioned in the present connection that W. Linnik,† in 1930, succeeded in photographing the interference fringes which arise due to the reflection of X-rays by the method of Lloyd's single mirror. He worked in succession with wave-lengths of about 1.6 and 2.0 Å. Especial technique had to be devised to render the experiment successful. The fringes are so narrow that a magnification of about 200 is required to render them conveniently distinguishable.

24. The Double Crystal Spectrometer.‡—Referring to Figure 227, consider the plane of the drawing to be horizontal and let T represent the target of an X-ray tube, S_1 the entrance slit of the spectrometer, A the first crystal, S_2 a second slit, B the second crystal, supposedly similar to the first one, and C an ionization chamber. The slit S_1 is rather wide, perhaps 1.5 mm.

* *Spektroskopie der Roentgenstrahlen* (2d ed.), pp. 51 ff. † *Zeit. f. Phys.*, **65**, 107, 1930.

‡ Selected references:

1. A. H. Compton, *Phys. Rev.*, **10**, 95, 1917.
2. W. L. Bragg, James, and Bosanquet, *Phil. Mag.*, **41**, 309; **42**, 1, 1921.
3. Davis and Stempel, *Phys. Rev.*, **17**, 608, 1921.
4. Davis and Stempel, *ibid.*, **19**, 504, 1922.
5. Wagner and Kulenkampff, *Annalen*, **68**, 369, 1922.
6. Davis and Terrill, *Phil. Mag.*, **45**, 463, 1923.
7. B. Davis, *Proc. Nat. Acad.*, **13**, 419, 1927.
8. Ehrenberg and Mark, and Ehrenberg and Susich, *Zeit. f. Phys.*, **42**, 807 and 823, 1927.
9. M. Schwarzschild, *Phys. Rev.*, **32**, 162, 1928.
10. Davis and Purks, *ibid.*, **34**, 181, 1929.
11. S. K. Allison, *ibid.*, p. 176.
12. Allison and Williams, *ibid.*, **35**, 149 and 1476, 1930.
13. Du Mond and Hoyt, *ibid.*, **36**, 1702, 1930; **37**, 1443, 1931.
14. S. K. Allison, *ibid.*, **41**, 1, 1932.
15. Allison, *ibid.*, **44**, 63, 1933.
16. L. G. Parratt, *ibid.*, **41**, 553 and 561, 1932; **44**, 695, 1933.
17. Kirkpatrick and Ross, *ibid.*, **43**, 596, 1933.

For many purposes it is necessary to use two slits between the target and the first crystal instead of a single one. These then have equal width and are separated by a distance of several centimeters taken parallel to the beam. The slit S_2 is not essential; it is often omitted entirely. However, when present, it is wider than the slit, or slit system S_1. Furthermore, the window of the ionization chamber is perhaps twice as wide as S_1.

To insure similarity of the reflecting faces of the crystals A and B, two surfaces are used which were contiguous before cleaving an originally single crystal.

Let us consider the beam of parallel rays represented by the solid lines, issuing from various points of the target, and all incident upon the crystal A at the same glancing angle θ_1. The various parts of this beam will all reflect the same wave-length λ_1, determined by the law according to which Bragg reflection occurs; we shall not be concerned with the question of whether Bragg's law is to be used in its corrected or its original form, so we may simply write $m\lambda = 2D \sin \theta$. The reflected beam passes through the slit S_2 and is incident upon crystal B, which is supposedly so adjusted that the once reflected beam will make the same glancing angle θ_1 at crystal B as at A. Hence the beam λ_1 will be reflected a second time and then enter the ionization chamber C.

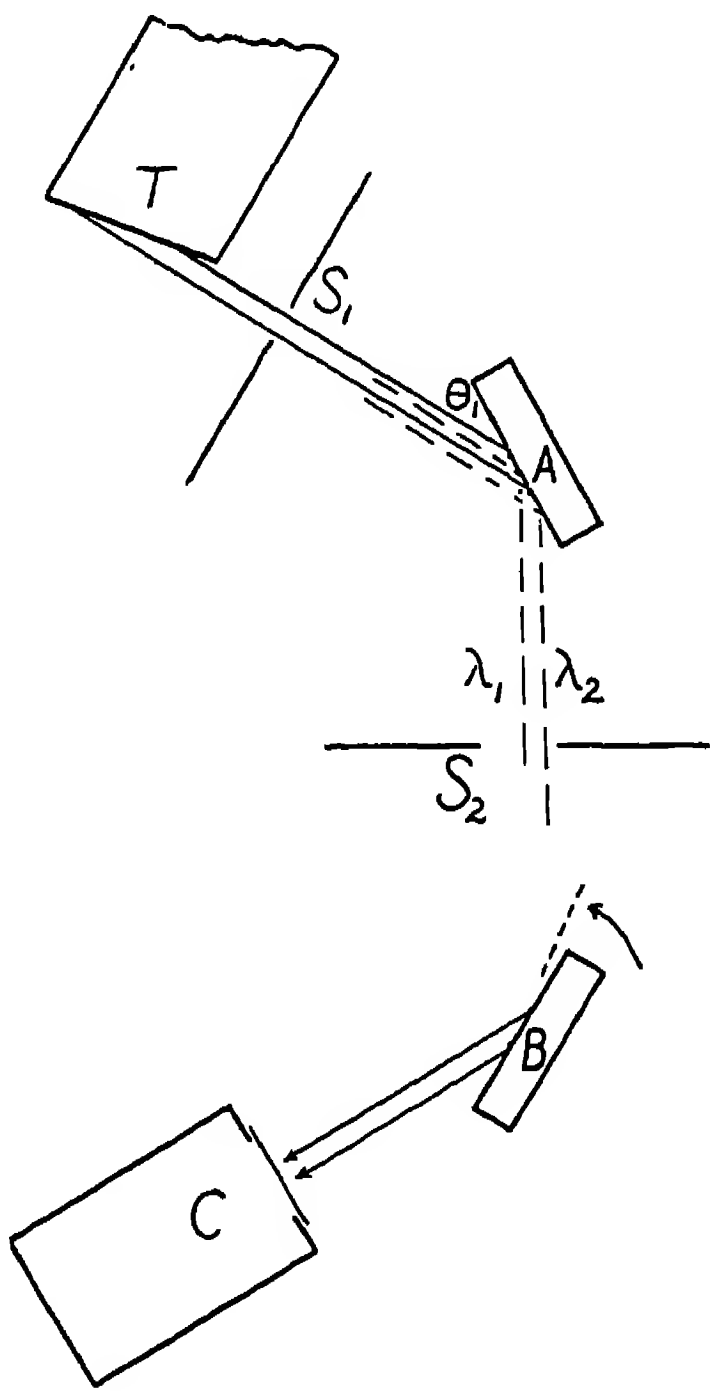


FIG. 227

The first crystal plays the rôle of a collimator in that it selects out of the original divergent beam, for presentation to the second crystal, only rays which are parallel, for each given wave-length within a small range of wave-lengths which the first crystal reflects because a small range of glancing angles of incidence obtains.

Now let us consider a different parallel beam—for example, the one represented by the broken lines. This makes a slightly smaller glancing angle at A; let us denote the glancing angle by θ_2. Out of this beam a slightly smaller wave-length λ_2 will be reflected by crystal A. The once reflected beam λ_2 will also pass through the slit S_2, but it will be incident upon crystal B at a glancing angle equal to $[\theta_1 + (\theta_1 - \theta_2)]$, that is, at an angle greater than θ_2,

in fact greater than θ_1, and hence the wave-length λ_2 will fail to be reflected a second time. However, if crystal B is "rocked" slightly, in the direction of the arrow, to let us say the position indicated, or precisely, through the angle $2(\theta_1-\theta_2)$, the glancing angle of the beam λ_2 upon B will assume the value θ_2, and hence the wave-length λ_2 will now be reflected and λ_1 will fail to be reflected. The wave-lengths λ_1 and λ_2 will thus be easily *resolved* if the difference $\Delta\lambda=(\lambda_1-\lambda_2)$ is not all too small. How small this difference may permissibly be will depend among other factors upon the magnitude of the small "angle of tolerance" over which each of the wave-lengths λ_1 and λ_2 is reflected by the crystals, and when the crystals are good ones this angle is very small indeed. Hence the resolving power of the double spectrometer is very high when good crystals are used—as the resolving power of X-ray spectrometers goes. Numerical values of $\lambda/\Delta\lambda=10{,}000$ have been obtained (No. 14 of the previous footnote). Further, if we consider the difference $(\lambda_1-\lambda_2)$ to be the lowest permissible for resolution by the double spectrometer, and consider a photographic plate to be erected in the plane S_2, the wave-lengths λ_1 and λ_2 will not be nearly resolved by the single spectrograph which is then operative; and while the resolving power of a single crystal spectrograph may be enhanced by narrowing the entrance slit, a very narrow slit must be used, for example, .02 mm, and the distance at which the photographic plate is erected must be several meters, in order to attain resolving power at all approaching that of the double spectrometer. The intensity is then very low, and, accordingly, exposures of let us say forty or fifty hours are necessary when investigations requiring high resolving power are undertaken with a single crystal spectrograph. When such investigations are undertaken the Seemann slit spectrograph is generally used.

The resolving power of the double spectrometer is, on the other hand, not a direct function of the slit width at all, because the angle through which the second crystal must be rotated from its original position to prevent reflection of the wave-length λ_1 is not a direct function of the slit width. However, widening the slit will cause an additional portion of each crystal to be effective, and if the crystals are poor or only moderately good, this portion may not be a uniform continuation of the portion originally effective; in this case somewhat of a lowering of the resolving power will result indirectly, when the slit is widened.

The fact that wide slits may be used with the double spectrometer means that this instrument yields great intensity.

The slits should, however, be diaphragmed in length or "height." If too great a height is used, the rays which are most inclined upward or downward will make glancing angles at the crystals which differ sufficiently from the angles made by the corresponding rays lying in the horizontal plane, to

spoil the otherwise high resolving power. This matter of diaphragming in height is consequently one of importance. The height which is permissible depends upon the distance between the slits. The square of the angle of maximum vertical divergence of the beam, expressed in radians, should be only of the order of magnitude of the tolerance angle of the crystal. In practice this usually leads to a permissible height of only some 2 or 3 mm.

The total angle through which the second crystal needs to be rocked is only a few minutes of arc. It is usual to rock the crystal through the necessary small angle in very small steps—in steps of a few seconds of arc only—and to determine the intensity of the twice reflected beam at each step. A "rocking curve" is then plotted with angular setting or its equivalent, wave-length, as abscissa, and with intensity as ordinate (see Fig. 228).

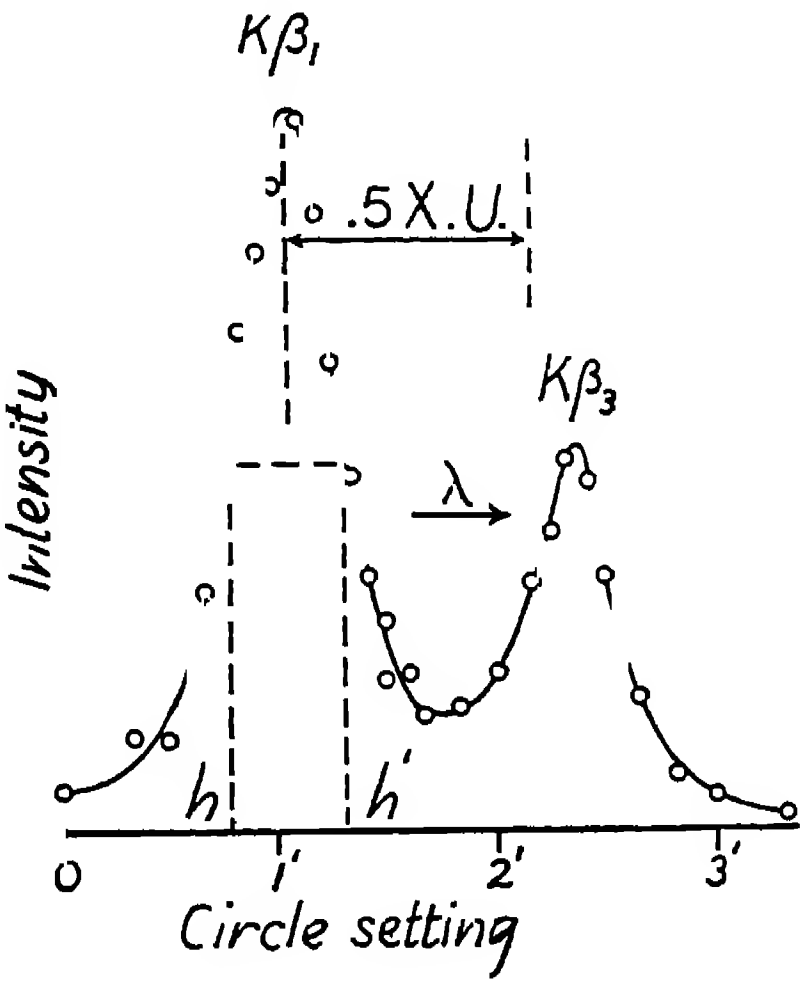


FIG. 228.—Rocking curve taken over the *Mo* $K\beta$ line, showing the doublet structure of this line (redrawn from Allison and Williams, *Phys. Rev.*, **35**, 149, 1930 [Fig. 1, curve *a*]).

With the double spectrometer the use of the ionization method is a necessity; the photographic plate and the ionization chamber cannot be used alternatively, as they can in the single spectrometer.

The field of usefulness of the double spectrometer, used in the manner so far described, is in investigations of the following kinds: In examining spectrum lines with a view to determining whether they may not really consist of close doublets; in measuring the separation of the components of a doublet; in measuring the relative intensities of the components of a doublet or of other neighboring spectrum lines; and, finally, in measuring the intrinsic spectral or "natural" widths of spectrum lines. By way of example: The Mo $K\beta$ line, $\lambda = 631$ X.U. has been separated into components, as represented in Figure 228. The components are designated as Mo $K\beta_1$ and Mo $K\beta_3$; the angle through which the crystal B must be turned to pass from the center of one component to the center of the other is 78 seconds of arc, when the crystals, of calcite, are oriented to reflect the wave-length 631 X.U. in the second order; the corresponding wave-length difference, deducible from Bragg's law, is only .57 X.U.! The intensity of $K\beta_3$ is evidently about one-half of that of $K\beta_1$. The width of the curve for the component

$K\beta_1$, measured, as is customary, where the curve has one-half of its maximum height, is represented by the distance hh', and is equal to 31 seconds of arc, or about .23 X.U. The width of the curve is measured at half-height rather than at the base, because the measurement at the base would involve great uncertainty. However, instead of stating the full width measured at half-height, hh', the half-width at half-height, $hh'/2$, is usually stated. This equals the horizontal distance from the maximum to either of the two points of half-height when the curve is symmetrical, as it usually is. But in whatever terms the width may be stated, this width, for the observed curve, is not all attributable to spectral or natural width of the spectrum line. A portion of the observed width has its origin in the instrument: The beam has a certain divergence in the vertical plane deducible from the known height of the slits and their distance of separation along the beam. This divergence carries with it the fact that a small range of glancing angle obtains in the vertical plane and hence a small range of wave-lengths is reflected at any one setting of the crystal. This range can be calculated and a correction applied when desired (see p. 348 n., Nos. 13, 14, and 15). Strictly speaking, a correction should also be applied for the small but finite angle of tolerance toward variation of glancing angle which the crystals exhibit, but when good crystals are used this correction is very small and is of doubtful value (Nos. 8 and 15). Interesting information regarding the actual amount of tolerance which various crystals exhibit toward variation of the glancing angle may be obtained, however, by means of the double crystal spectrometer, by using this instrument in a manner differing from that thus far considered:

Imagine the crystal B rotated about a vertical axis through the center of its reflecting face until this face becomes parallel with the reflecting face of crystal A; the spectrometer will then be in the so-called "parallel" position. The position previously considered is called "antiparallel." In the parallel position, not only the central beam, of wave-length λ_1, makes the same glancing angle θ_1, at both crystals, but any other beam which strikes both crystals, of wave-length λ_2, also makes the same glancing angle θ_2 at the second crystal which it makes at the first one. Moreover, rotation of the second crystal will now cause all beams incident upon this to be incident at an angle which does not correspond to their wave-length and will hence cause all beams to fail to be reflected. A small rotation of the second crystal will suffice to reduce the intensity of the beam entering the ionization chamber to zero, and the magnitude of the required rotation is determined by the angle of tolerance of the crystals. When in the instrument used in obtaining the curve of Figure 228, in the antiparallel position, the crystal B was turned to the parallel position, and a rocking curve was taken

in this position, the half-width of the curve at half-height was only 1.5 seconds of arc. In other words, a rotation of the second crystal of some three or four seconds of arc from the accurately parallel position sufficed to reduce the intensity to zero. (It is of interest to compare this value with the half-value of the angle of tolerance deducible from the theory given in section 21, remembering that the diffraction was of the second order; equality of the two values is, however, not to be expected.)

The width of the rocking curve obtained in the parallel position is *not* a direct function of the magnitude of the small range of wave-lengths which the ionization chamber admits when the second crystal is accurately parallel to the first one, because all of the wave-lengths falling into this small range will be simultaneously extinguished when the second crystal is rotated. Consequently, in the parallel position, rotation of the second crystal gives no information whatever regarding the intensity distribution which obtains in the small range of the spectrum which is effective. A small portion of the continuous spectrum, with or without one or more spectrum lines present within its limits, leads to practically the same width of rocking curve, taken always at half of maximum height, as a single narrow spectrum line alone present. In brief, in the parallel position, rocking of the second crystal gives no information whatever concerning the spectrum of the source, but does give valuable information regarding the nature of the reflection produced by the crystals.

In the parallel position, as in the antiparallel, changing the slit width does not affect the width of the rocking curve directly, but may affect it indirectly. Moreover, in the parallel position, changing the effective height of slit has no direct effect, because the width of the spectral range which is effective is of no concern, whereas in the antiparallel position, too great effective slit height has a direct effect of reducing resolving power.

The half-width at half-height of the rocking curve of 1.5 seconds of arc, in the parallel position, which was mentioned above, was observed with a pair of very good calcite crystals reflecting in the second order. When reflecting in the first order, the half-width at half-height would have been some 3 or 4 seconds of arc. Less perfect calcite crystals were found, in one case (p. 348 n., No. 11), to give, in the first order, a rocking curve having a half-width at half-height of 12 seconds, showing that appreciable disorientation of microcrystals existed. And crystals of rocksalt, which are much less perfect than calcite, give half-widths at half-heights of 5 or 10 *minutes* of arc (p. 348 n., No. 4). Imperfect crystals give broader and lower rocking curves than more perfect ones. Crystals giving wide rocking curves should not be used in the double spectrometer for the resolution of doublets (in the antiparallel position). The information gained by studying

the reflection of less perfect crystals, in the parallel position, is, however, of interest for itself. In fact, the double spectrometer was originally used only in the parallel position for the purpose of investigating the nature of crystal reflection:

In such investigations since two originally contiguous faces are used, it is reasonable to assume that the two faces are alike in the degree of disorientation of the microcrystals, and hence the observed rocking curve may be considered to be the sum of two equal components, one contributed by each crystal. The observed curve has somewhat the form of a Gaussian error distribution. According to theory applying to Gaussian error curves, the width of each of two equal component curves, at half-height, is equal to the width of the observed curve, at half-height, divided by $\sqrt{2}$. The width of the observed curve, at half-height, divided by $\sqrt{2}$ is then the width, at half-height, of the rocking curve attributed to each crystal separately.

When the crystal faces are each prepared by similar grinding and polishing after cleavage, a half-width of the observed rocking curve, at half-height, for two calcite crystals, of as much as 30 seconds of arc, has been observed, in the first order (p. 348 n., Nos. 3 and 5). This shows that additional disorientation is introduced near the crystal surfaces by grinding and polishing.

However, not all of the width of the rocking curve is to be attributed to disorientation of the microcrystals in any case. Even perfect crystals would yield rocking curves of finite width because of the small angle of tolerance toward variation of the glancing angle while reflecting a monochromatic beam which even perfect crystals would exhibit.

Recent careful investigation has shown that the best calcite crystals, used without preparation of any kind, yield rocking curves having only the width expected for ideal crystals, indicating that practically no disorientation exists (p. 348 n., No. 14).

The double spectrometer is also used in the parallel position for determining the amount of energy of a given wave-length which given crystals reflect. Two different quantities are determined, namely, the "percentage of reflection" and the "coefficient of reflection." Each of these two quantities is defined in its own distinct manner:

In determining percentage of reflection the intensity of the beam which is to be incident upon the second crystal is first determined by removing the second crystal from the spectrometer and turning the ionization chamber to the position in which the beam from the first crystal will enter the chamber directly. The window of the chamber must be large enough to take in the entire beam from crystal A, or, if there is a slit S_2 present, then the window must admit at least all of the beam which passes through this slit. The in-

tensity is measured by the number of scale divisions of deflection of the electrometer per second of time. After this measurement has been made, the ionization chamber is turned to the location in which it will receive the twice-reflected beam, in the parallel position, and the second crystal is inserted and adjusted, and set at the circle reading of the maximum height of the rocking curve. The deflection of the electrometer per second of time is again measured, and the quotient of the two intensities times 100 is the percentage of reflection. Pairs of calcite crystals of high grade have been found to give, in the first order, from 30 to 35 per cent of reflection at about $\lambda =$ 200 X.U., and increasing reflection with increasing wave-length up to 60 or 65 per cent at about $\lambda = 1200$ X.U. At greater wave-lengths the reflection was found to fall off slowly (p. 348 n., No. 14). Prepared faces of calcite give percentages which are perhaps one-half as great as for unprepared faces. Rocksalt, either with or without preparation, reflects, in the first order, from 5 to 20 per cent, depending upon the specimen, the wave-length region, etc. (p. 348 n., Nos. 4 and 17).

It is important that the two crystals be similar, in determining percentage reflection, for a change in the degree of perfection of crystal A would change the amount of divergence which exists in the beam of any given wave-length which is incident upon B. And a change in this divergence will alter the percentage of the given wave-length which crystal B reflects. That is, the recorded percentage is determined by the two crystals jointly and is characteristic of each separately only when it is characteristic of both at the same time.

While percentage of reflection is lowered by disorientation of the microcrystals, the method of calculating "coefficient of reflection," which we shall now discuss, gives the disoriented microcrystals an opportunity to contribute to the result.

The "coefficient of reflection," to be denoted by R, may be defined as the area under the rocking curve divided by the intensity of the incident beam, I. The intensity I is measured just as when determining percentage, namely, by removing the second crystal and turning the ionization chamber to the position in which it will receive the beam from crystal A directly—and then observing the number of scale divisions of deflection of the electrometer per second of time. The ordinates of the rocking curve are now to be plotted in terms of deflection of the electrometer per second of time at each angular setting of the second crystal, and the abscissae of the rocking curve are to be plotted in some chosen unit of angle which must be stated in connection with the final result. The unit chosen is usually the radian, or the degree. For a pair of excellent calcite crystals, values of R have been recorded which increase not very far from linearly with wave-length from

about 1.2×10^{-5} at $\lambda=200$ X.U. to about 4.7×10^{-5} at $\lambda=2300$ X.U. (p. 348 n., No. 14), the unit of angle being the radian. To express R in terms of the degree as the unit of angle of the abscissa, the foregoing numbers must each be multiplied by 57.3. Crystals having a great amount of disorientation in their structure give much *higher* values of R. For example, with rocksalt, values of R of 8.4×10^{-4} at wave-length about 300 X.U., and 3.7×10^{-4} at wave-length about 700 X.U., have been found (p. 348 n., No. 6), the unit of abscissa again being the radian. With rocksalt the value of R *decreases* with increase of wave-length. (For data on extraordinarily good specimens of rocksalt see p. 348 n., No. 17.)

A definition of coefficient of reflection is often given which differs in form from that given above but is equivalent to it. This definition requires explanatory discussion: According to the definition given above, R equals the area under the rocking curve divided by I, the intensity of the incident beam. Consider the axis of abscissae of the rocking curve to be divided into a large number of equal small elements $\Delta\theta$. Draw an ordinate of the curve at the center of each division, and denote this by e; the value of e will, of course, be different for each division. The area under the rocking curve is $\Sigma e\Delta\theta$, or, since $\Delta\theta$ is the same for each element, the area is $\Delta\theta\Sigma e$, that is, simply the sum of the ordinates times $\Delta\theta$. Denote the sum Σe by E_s, then $R=E_s\Delta\theta/I$. Now E_s is the sum of the number of scale divisions of deflection of the electrometer which would be obtained if the crystal were set for one second at the center of each interval $\Delta\theta$, or is, in other words, the total or accumulated deflection of the electrometer which would be obtained. But this total deflection is the same as the total deflection of the electrometer which would accumulate if the crystal were rotated continuously across the extent of the entire rocking curve with an angular velocity of $\Delta\theta$ per second. Further, if this angular velocity were doubled, becoming $\omega=2\Delta\theta$ per second, the deflection which would then accumulate would be $E=\frac{1}{2}E_s$. That is, the product of the angular velocity and the corresponding accumulated deflection of the electrometer is constant, or $E_s\Delta\theta=E\omega$. The coefficient of reflection thus becomes:

$$R=\frac{\text{Area}}{I}=\frac{\Sigma e\Delta\theta}{I}=\frac{E_s\Delta\theta}{I}=\frac{E\omega}{I}. \tag{31}$$

The coefficient of reflection may thus be alternatively defined as the product of E, the total deflection of the electrometer which accumulates while the second crystal is rotated uniformly over a region including all of the rocking curve, and the angular velocity of the rotation ω, this product being then divided by the deflection of the electrometer which is obtained per second when the ionization chamber receives the beam from crystal A directly.

The unit of angle in which the speed of rotation is measured must always be stated—and this is the only unit which enters into the final result; if the basic time unit were altered, ω and I would change by equal factors.

The measured value of the coefficient of reflection is not greatly affected by substitution for the first crystal of a specimen of a different degree of perfection. Such substitution would alter the intensity of the incident beam, and would alter both the shape and the area of the rocking curve. However, the area would be altered in roughly the same proportion as the intensity of the incident beam, leaving the coefficient approximately unaltered.

The value of the reflection coefficient is affected by the state of polarization of the beam, but this is a matter which cannot be entered into here.

We have thus far considered only the cases, parallel and antiparallel, in which both crystals reflect in the same spectral order. While these cases comprise the most general and most important use of the double spectrometer, there is no reason why the crystal A should not be set to reflect in, for example, the first order and crystal B in the second order. For the purpose of designating the various possible combinations of orders a special notation is used: Let m_A be the order in which the first crystal reflects, this being always considered numerically positive. Whenever, as in Figure 227, the twice-reflected beam, BC, lies on the same side of the beam AB as the original beam, TA, the order m_B in which the second crystal reflects is to be considered positive, and when the twice-reflected beam lies on the opposite side of BA, as in the parallel position, m_B is to be given a negative sign. Thus, for example, when both crystals reflect in the second order and the crystals are in the parallel position, the order for the instrument as a whole is (2, −2), and when in the antiparallel position, the order is (2, 2). The foregoing is referred to as "Allison's notation"— it was first suggested in a paper by Allison and Williams.*

A few words should be said concerning the dispersion which the instrument yields: According to Bragg's law, $m\lambda = 2D \sin \theta$. Differentiating, we have for the angular dispersion of the beam leaving the first crystal $d\theta_A/d\lambda = m_A/2D \cos \theta_A$. At crystal B the additional dispersion $d\theta_B/d\lambda = m_B/2D \cos \theta_B$ is introduced. The total dispersion is the sum of these, or:

$$\frac{d\theta_t}{d\lambda} = \frac{m_A}{2D \cos \theta_A} + \frac{m_B}{2D \cos \theta_B}, \tag{32}$$

and this is all recorded by rotating crystal B. The dispersion with both crystals reflecting in the same order in the antiparallel position ($m_B = m_A$),

* *Phys. Rev.*, **35**, 149, 1930. The original statement of the rule to be followed in assigning the order was of less general applicability than the statement of the rule now generally made, which we have followed.

is thus twice that due to a single crystal reflecting in the same order. And in the parallel position ($m_B = -m_A$), the dispersion is zero. When m_B is negative and numerically greater than m_A, the dispersion is negative.

The double spectrometer was first used in the year 1917 by A. H. Compton for determining the coefficient of reflection of one specimen of calcite, and of one of rocksalt, but only a very incomplete report of the work was published. In the year 1921 W. L. Bragg, James, and Bosanquet, measured the coefficient of reflection for what was probably a rather poor specimen of rocksalt, and in the same year Davis and Stempel published rocking curves for three pairs of calcite crystals and measured the percentage of reflection at various wave-lengths. In 1922 these authors published similar results for rocksalt. In the same year Wagner and Kulenkampff contributed some further data on calcite and on rocksalt, and in a detailed footnote discussed the possibility of using the double spectrometer in the antiparallel position for the resolution of close doublets. They, however, came to the conclusion that the instrument could not be expected to resolve close doublets because of the lowering of resolving power caused by rays which are rather steeply inclined upward or downward with reference to the horizontal plane. For some reason they failed to conceive the idea of introducing diaphragms to limit the vertical aperture of the beam! After five more years had elapsed, in 1927, Bergen Davis, and independently Ehrenberg and Mark, and Ehrenberg and Susich, reported the first use of the double spectrometer in the antiparallel position for the resolution of doublets and the determination of the natural widths of spectrum lines. These investigators used beams of small vertical aperture. Thereupon—that is, as soon as the possibilities of the double spectrometer in the antiparallel position were once fully realized—the instrument gained rapidly in importance—and it promises to continue as one of the outstanding instruments of X-ray research. Both the theory of the instrument and the technique have been developed of late years especially by S. K. Allison, beginning in the year 1929.

Du Mond and Hoyt have pointed out certain objections to the usual procedure of rocking only the second crystal when the instrument is used in the antiparallel position for measuring the natural widths or relative intensities of spectrum lines. The principal objection is that during the progress of the rotation the beam which is effective wanders across the faces of both crystals and the face of the target. This wandering introduces slight fortuitous changes in intensity, and hence, to avoid this wandering, Du Mond and Hoyt constructed a spectrometer which rotates both crystals simultaneously in opposite directions, and rotates the remaining parts of the apparatus correspondingly as need be. This plan is being adopted also by others.

25. Liquids, Amorphous Solids, and Gases.—Little more can be done in this book than to call attention to the fact that extensive investigations of the diffraction effects which are produced by liquids, by amorphous solids, and by gases have been carried out; the methods of attack and the nature of the results obtained will, however, be briefly outlined.

Considering *liquids* first: The liquid is usually placed in a small cell having very thin walls of paraffined paper, mica, or glass, or it may be suspended at the end of a rod in the form of a drop. The specimen is to be thought of as replacing the crystal powder in the powder spectrograph (Fig. 207), being placed at C. Monochromatic or practically monochromatic X-rays are used, just as when the instrument in question is used for its original purpose of investigating a crystal powder. With liquids, broad halos are found, in place of narrow lines. The halos, like the lines, surround the spot formed by the undeviated beam. Direct intensity measurements may be made by replacing the photographic plate or film by an ionization chamber and rotating this about the point C step by step.

For many liquids only a single halo is found. A mathematical theory has been developed which explains the existence of this halo on the basis of interference between wave trains scattered by symmetrical molecules, of course without orientation.* The "average spacing" between molecules determines the position of the halo. It develops that Bragg's law, or a slight modification of it, can be applied. For some liquids a principal halo and one or two fainter halos are found and the latter are not accounted for by the aforementioned theory.

In order to account for the observed diffraction effects, the molecules of many liquids must be considered to form small groups. Much of the evidence that these small groups exist in many liquids has been supplied by the work of Stewart and his collaborators,† who consider that the liquid state resembles the crystalline state more nearly than it does the gaseous state, which will be discussed presently. They term the condition which obtains in liquids *cybotaxis*, meaning "space arrangement." It is of wide occurrence. The molecules in the groups have regularity of position and orientation, but the groups have only a transient existence and do not approach in size, and stability, those of microcrystals in solids. This theory is supported by experiments upon several chemical series, particularly those composed of unsymmetrical molecules of the "chain" type whose space grouping is made very reasonable by the lack of symmetry. Two or three halos occur in the patterns. By applying Bragg's law, consistent values for the lengths and

* Raman and Ramanathan, *Proc. Ind. Assoc. for Culti. of Sci.*, **8**, 127, 1923.

† See, e.g., G. W. Stewart, *Phys. Rev.*, **32**, 558, 1928; *Rev. Mod. Phys.*, **2**, 116, 1930; and *Phys. Rev.*, **35**, 726, 1930.

diameters of the molecules are obtained. It is also possible to determine, for example, the length occupied by each CH_2 group along a chain. This equals the difference in length of two kinds of molecules when the longer molecules differ from the shorter only in containing one additional CH_2 group. The foregoing refers to so-called "amorphous" liquids.*

In a class of substances known as "liquid crystals" there is independent evidence of the orientation of molecules within aggregates. A considerable number of organic compounds are known, notably *p*-azoxyanisole and *p*-azoxyphenetole, which upon melting become turbid liquids.† When these compounds are heated above the melting point, the turbidity remains until a definite, higher temperature is reached, whereupon they become clear. The liquids in the turbid condition thus represent a distinct state of aggregation or *phase*. Moreover, when they are in this condition they are optically anisotropic, showing double refraction. Minute aggregates are discernible when the liquid is examined under the microscope between crossed nicols. The molecules of these compounds are all of the "long chain" type. Straight chains are more favorable than chains which have a bend in them.‡ It is probable that in the liquid-crystal state the molecules are parallel within aggregates, and that this is the reason why the liquids are optically anisotropic. The turbid appearance is attributed to differences in orientation of aggregates, light being scattered at the intervening boundaries. The diffraction halos from liquid crystals are similar to those from ordinary or amorphous liquids.

The patterns formed by *amorphous solids*, for example, the resins, unstretched rubber, and glass, resemble those from liquids.§ Amorphous solids have been the least extensively studied.

The diffraction or scattering which is produced by *gases* is studied by methods which are basically those of the powder spectrograph and hence those which are applied also to the study of diffraction due to amorphous solids and liquids. But with gases the experimental arrangement must be modified in such a manner as to cause the X-ray beam to irradiate a greater volume of the material. We shall not enter into the details of how this is accomplished.

* For further information and detailed references, see the first two articles of the previous footnote, which are summaries, and also C. Drucker, *Phys. Zeit.*, **29**, 273, 1929, and J. A.Prins, *Zeit. f. Phys.*, **56**, 617, 1929.

† See, e.g., A. Findlay, "The Phase Rule and Its Applications" (5th ed., Longmans, Green & Co., 1923), p. 40.

‡ D. Vorlaender, *Ber. d. Deutsch. Chem. Gesell.*, **62**, 2824, 1929.

§ See, e.g., Parmelee, Clark, and Badger, *Jour. Soc. of Glass Technology*, **13**, 285, 1929; Clark and Amber, *ibid.*, p. 290; P. Krishnamurti, *Indian Jour. of Phys.*, **4** (2), 99, 1929; Randall, Rooksby, and Cooper, *Nature*, **129**, 458, 1930.

In a gas, since the molecules have a separation which is not only on the average much greater than in a liquid but is rapidly changing over a vastly greater range of value than in a liquid, the diffraction pattern is *not* produced primarily by interference among wave trains scattered by the *molecules* considered as centers of scattering, but chiefly by the *atoms* within one molecule considered as centers of scattering. The scattering is not *inter*-molecular but is *intra*molecular or *inter*atomic.

(With liquids or amorphous solids, interatomic interference effects are supposedly also present in the patterns which are obtained, but these effects are forced into the background by the effects produced by intermolecular interference, which dominate in the formation of the pattern.)

Interesting examples of the patterns formed by *polyatomic* gases are those formed by the two isomeric forms of dichloro-ethylene ($C_2H_2Cl_2$) in gaseous form, which have been studied photographically.* In the so-called "trans" form of $C_2H_2Cl_2$ the two Cl atoms lie at two diagonally opposite corners of a rectangle and the two much lighter hydrogen atoms, which contribute little to the scattering, lie at the two remaining corners. The carbon atoms lie along the long central line of the rectangle in both the "trans" form and, in the other, the "cis" form. In the "cis" form, however, the two Cl atoms lie at corners on the same side of the rectangle, and are thus closer together than in the "trans" form. The two H atoms lie at the two corners on the other side of the rectangle, but these hardly count, because they are so light. The Cl atoms, which dominate the scattering, are thus further apart in the "trans" form than in the "cis" form, and consequently in the pattern produced by the "trans" form the halo is of smaller radius than in that produced by the "cis" form. The X-ray diffraction patterns thus furnish a very valuable check upon the chemical evidence regarding the position of the chlorine substitution in these two isomers.

In the scattering produced by *monatomic* gases there is no possibility of interatomic interference, but intra-atomic or inter-electronic interference will yet occur. The question of atom structure factor enters here. The wave trains scattered by the electrons surrounding the nucleus of any one atom interfere among themselves, but this interference does not produce a defined maximum in the scattering. Measurement of the scattering produced by monatomic gases is of interest and importance because it yields a direct determination of the value of the atom-structure factor as a function of the angle of scattering.

For further information regarding the diffraction or scattering of X-rays by gases see the references given below.†

* P. Debye, *Phys. Zeit.*, **31**, 142, 1930.

† E. O. Wollan, *Rev. Mod. Phys.*, **4**, 205, 1932; H. Gajewski, *Phys. Zeit.*, **33**, 122, 1932.

CHAPTER 9

THE DIFFRACTION OF MATERIAL PARTICLES

When a stream of electrons bombards a crystal, one or more strong and well-defined streams of electrons may issue from the bombarded area. This startling fact was discovered by C. J. Davisson and L. H. Germer of the Bell Telephone Laboratories, in a manner of which we shall tell later. The experiment must be performed in a high vacuum. The incident stream is produced by a so-called "electron gun," and the issuing streams are received by a "Faraday box collector," both of which will be described in due course. Referring to Figure 229, let the incident stream of electrons be i, and let the crystal, the "target," be T. When the incident electrons all have the same, suitable velocity, there will be a single issuing beam such as represented by the solid curve in the figure. This is a polar graph, that is, the number of electrons issuing from the point of bombardment in any direction is proportional to the length of the vector v drawn from the point of bombardment in the given direction. The issuing beam calls to mind a beam of light diffracted by a grating, or a beam of X-rays diffracted by a crystal. The occurrence of such electron beams forces us to entertain the hypothesis that electrons have *waves* in some manner associated with them and that these waves are being diffracted by the crystal lattice. The surface layer of atoms may be acting as a grating, or several layers of atoms may be acting as a space lattice.

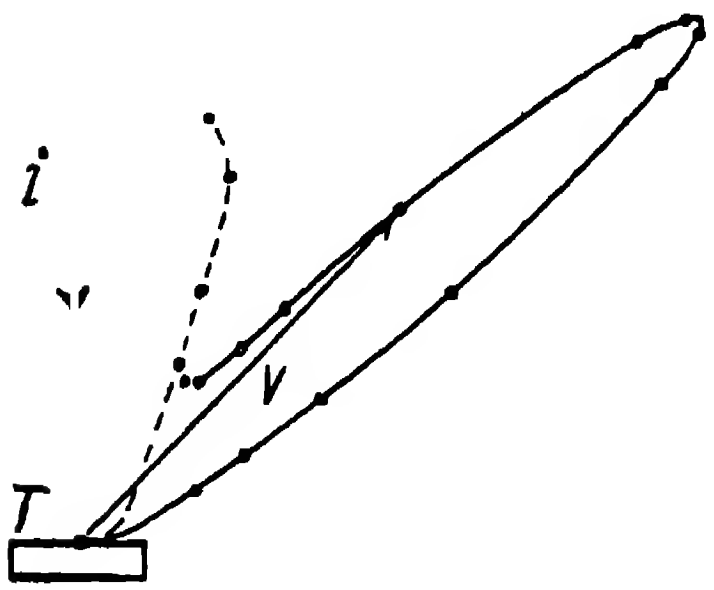


FIG. 229.—Solid curve, beam of electrons issuing from bombarded area of a nickel crystal (Davisson and Germer). Broken curve, background scattering.

When we try to account for the strong and well-defined issuing beams upon the supposition that electrons have only the nature of particles, our failure is complete.

If we regard the atoms of the crystal as "solid," the bombarding of the crystal by electrons is to be likened to the shooting of birdshot at a pile of cannon balls in regular array. The birdshot would not rebound in a well-defined beam but would scatter nearly equally over a large range of directions.

Again, we may regard each atom of the crystal—nucleus and surrounding electrons—as a "solar system," and might regard each incident electron as a "comet." If now the original path of an incident electron passes sufficiently close to the nucleus of an atom, the passing electron might be attracted by the nucleus with sufficient force to cause a deflection in path of over 90°, but there would be no strongly preferred direction for the issuing electrons. Moreover, neighboring atoms would be without effect and hence the state of aggregation of the atoms of the target should be immaterial. But this is contrary to experiment. While a target consisting of a single crystal yields strong beams, one consisting of a multitude of small crystals yields no beams under the same conditions.

It might also be argued that the incident electrons are deflected by atoms which lie fairly well below the crystal surface and then escape most readily in those directions in which the passageways in the crystal are large, or, the rows of atoms are far apart, the so-called "transparent" directions of the crystal. But the directions taken by the issuing beams are found to be unrelated to the transparent directions, and therefore this hypothesis must be abandoned.

1. Davisson and Germer's Discovery.*—In April, 1925, Davisson and Germer were investigating the scattering of electrons by a target of ordinary, polycrystalline nickel, when, through a "fortunate accident," the target became fouled, and in order to clean the target they heated it to a high temperature while it was hidden in the apparatus. The heating caused the target to recrystallize into comparatively few larger crystals, but this fact was at first unknown to the experimenters. However, the scattering curve which they subsequently obtained differed from the curve previously obtained in showing several fairly well-defined lobes, which meant that there were now several fairly well-defined beams of electrons issuing from the target. Upon removing the target from the apparatus they discovered the recrystallization. Thereupon they turned to experimenting with a single crystal, and through a long and arduous series of experiments made their important discovery. However, had the "fortunate accident" of the fouling of the crystal not occurred, the discovery would hardly have been long delayed. Davisson and Germer had developed all of the elaborate and difficult technique which is necessary for the experiment, and the suggestion to try a target consisting of a single crystal was all that was needed. The suggestion would have been given by a theory which had just been advanced, but knowledge of this theory had not yet become disseminated when the discovery was made.

* *Nature*, **119**, 558, April, 1927; *Phys. Rev.*, **30**, 705, 1927; *Bell Syst. Tech. Jour.*, **7**, 106, 1928; *Jour. Franklin Inst.*, **205**, 597, 1928; *Jour. Chem. Educ.*, **5**, 1041 and 1255, 1928.

2. The Wave Theory of Matter.—In the year 1924 L. de Broglie in a Doctor's thesis advanced the daring hypothesis that matter has an intrinsic wave nature in addition to its nature of particles or corpuscles.* Previously matter had never been thought of as being other than corpuscular, whereas light had of old been thought of as being due either to corpuscles or to wave motion, and in recent decades it has been thought of as being due to both at the same time, in some manner of which we cannot, however, form a simple and clear picture. Convincing evidence for the wave nature of light was adduced during the last century, and equally good evidence that light has a corpuscular nature, because it is emitted and received in quanta, has been adduced during the present century. Thus, light evidently has a dual nature. And the same applies to other types of electromagnetic radiation, being especially well established also for X-rays. Hence, in view of the intimate relation which exists between matter and radiation, may not matter, too, have a dual nature of waves and corpuscles, the corpuscular nature being revealed to us in common experience, but the wave nature having hitherto remained unrevealed?

De Broglie formulated the foregoing brilliant idea which he had into a definite theory: The fundamental quantum principle for radiation is the relation between the energy E of a corpuscle of light, and the associated frequency, ν, namely, $E=h\nu$, where h is Planck's constant. The basis of De Broglie's theory is that this principle may be extended to material particles. Accordingly, a particle having energy E has waves of frequency ν in some manner "associated" with it and:

$$E=h\nu\,. \tag{1}$$

What these waves *are* we cannot at present say and shall perhaps never be able to say. They are at the basis of matter which is the most basic manifestation of nature of which we have hitherto claimed knowledge. May we ever hope to explain a thing which explains matter? The waves are referred to simply as "ψ-waves." It can be shown in a manner which will be indicated presently that they have a wave-length:

$$\lambda=\frac{h}{mv}\,, \tag{2}$$

where m is the mass and v the velocity of the particle. The wave-length is all we shall need to know in order to understand the diffraction effects produced by material particles, because these effects depend only upon the wave-length of the waves. As in light, the wave fronts are perpendicular to

* *Thèse* (Paris: Mason & Cie, 1924), or *Ann. de phys.*, **3**, 22, 1925.

the rays. The ray is the direction of propagation of the energy, or, for material particles, the direction of motion of the particle. For a particle in uniform motion the wave fronts are to be regarded as plane. When particles of a given kind all have the same velocity, the stream is said to be homogeneous. We shall be concerned mainly with homogeneous streams. They are equivalent to *plane monochromatic* waves. A heterogeneous stream in which the particles have a wide range of velocities is equivalent to white light and is referred to as "white radiation."

We shall now derive the expression for the wave-length, and discuss the relation of the waves to the particles (this derivation and discussion are submitted for the sake of completeness; understanding of them is *not* prerequisite for continuing with the main subject; statements and equations from the theory of relativity must of necessity be introduced and proof for these cannot be given here):

The mass of a particle is a function of its velocity. The energy of a particle is $E=mc^2$, where c is the limiting value for the transmission of energy or, practically, the velocity of light. Accordingly by equation (1):

$$\nu=\frac{mc^2}{h}. \tag{3}$$

The velocity of the ψ-waves may be shown to be:

$$V=\frac{c^2}{v}, \tag{4}$$

whence the wave-length is:

$$\lambda=\frac{V}{\nu}=\frac{h}{mv}, \tag{2'}$$

which is equation (2). Substituting the value of v into equation (4):

$$V=\frac{c^2m\lambda}{h}. \tag{5}$$

The velocity of the ψ-waves is a function of the wave-length.* The waves are dispersed even in empty space. In this dispersion they resemble the phase waves of light traveling in a transparent medium. The ψ-waves are known as "De Broglie phase waves." Further, the waves associated with a homogeneous stream of particles are monochromatic in the sense that a spectrum line is monochromatic—they comprise a narrow range of frequency.

* This is not a direct proportionality because m is variable. It is further to be noted that λ cannot be zero since v cannot exceed c.

Since the waves are dispersed, there will be a group velocity, U, differing from the phase velocity, V. Let us derive the value for U: If we call m_0 the mass of the particle at zero velocity, the so-called rest mass, then:

$$(mc)^2-(mv)^2=(m_0c)^2=\text{Const.} \tag{6}$$

Differentiating:

$$\frac{d(mc)}{d(mv)}=\frac{mv}{mc}. \tag{7}$$

The group velocity is, by equation (23) of chapter 7:

$$U=V-\lambda\frac{dV}{d\lambda},$$

and hence, by our present equations (2′), (3), and (7):

$$U=\frac{d(\nu)}{d\left(\frac{1}{\lambda}\right)}=\frac{\frac{c}{h}\,d(mc)}{\frac{d(mv)}{h}}=\frac{cmv}{mc}=v. \tag{8}$$

That is, the group velocity U is equal to the velocity of the particle v. The wave *group* carries the energy with it. It *is* the particle. The phase waves travel faster than the group—in fact, according to equation (4), faster than the velocity of light, since v cannot exceed c. The ψ-waves guide the particle just as the phase waves of light guide the light corpuscle. In the full development of the theory the phase waves of light and the ψ-waves are considered as different manifestations of the same thing.

The theory of De Broglie has been developed by Einstein, Scrödinger, Heisenberg, and others into the extensive subject, the "Wave Mechanics." Einstein's first contribution* came early in 1925 and this in turn led W. Elsasser,† in August of the same year, to state that it might be possible to find direct evidence for the existence of material waves by allowing slowly moving electrons to bombard a crystal. Low speed and small mass should mean relatively great wave-length. Hence the electrons might be observably *diffracted*. However, as we have learned, the diffraction in question was actually discovered independently of theoretical prediction.

When an electron is diffracted by a crystal, the wave fronts of the material waves must extend over at least several atoms of the crystal. This fact

* *Sitzungsber. d. Preuss. Akad. d. Wiss.* (1925), p. 1.

† *Naturwiss.*, **13**, 711, 1925.

sets a lower limit for the width of the beam. Further, by analogy with light and X-rays we must expect that an upper limit for the width of the beam will be roughly set by the aperture through which the beam passes. Between these wide limits we shall attempt to say nothing. The question of the number of wave crests in a group or "packet" will also be left unanswered except for the mention of occasional experimental data which bear upon it. Suffice it to say that we must be willing to regard electrons as waves under some circumstances, while still regarding them as particles under others. In doing this we are doing only what we do in the case of light and X-rays. As for the size which was formerly assigned to electrons, it was arrived at by extrapolating laws observed in gross experiments to micro-particles—an uncertain procedure. There is probably no definite size, and for an unbound electron the effective size, if we may use this term, may be much greater than that of an electron bound in an atom.

THE DIFFRACTION OF SLOWLY MOVING ELECTRONS
(Secs. 3–11)

3. Apparatus and Experimental Principles.—Referring to Figure 230, let G be an electron "gun," T the "target," and C a "collector." The collector may be rotated about an axis lying in the face of the target perpendicular to the plane of the figure. The entire assemblage is surrounded by a metal shield and the whole inclosed in a glass bulb and highly evacuated.

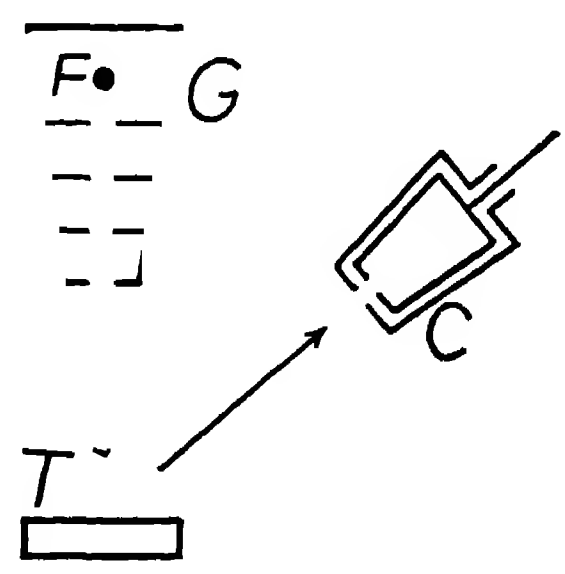


FIG. 230.—Electron gun, target, and collector.

The principle of operation of the gun is as follows: F is a tungsten filament which is heated by a current in a local circuit not shown. It serves as a source of electrons. The filament is at a negative potential relative to the wall of the gun. The electric field between filament and wall accelerates the electrons and some of these pass through a series of apertures and then pass from the muzzle of the gun. An expression will be found for the velocity, v, which the electrons acquire due to the field: They emerge from the filament with a velocity which is so small that they may be considered as starting from rest—or, in other words, with zero kinetic energy. The work done by the field on one electron is consequently equal to its total kinetic energy, or $W=\frac{1}{2}mv^2$. Potential difference, V, is work per unit charge, consequently:

$$V=\frac{W}{e}=\frac{mv^2}{2e},$$

where e is the charge of the electron. Or:

$$v = \sqrt{2\frac{e}{m}V}\,, \tag{9}$$

$$v = \sqrt{2(1.77\times10^7)(V\times10^8)} = 5.95\times10^7\times V^{\frac{1}{2}}\,. \tag{9'}$$

In the last equation v is measured in centimeters per second, and V in volts potential difference between the filament and the wall of the gun. Those electrons which escape through the first aperture in the gun find themselves in a field-free inclosure and neither gain nor lose velocity. The succession of perforated diaphragms serves to limit the beam and eliminate stray electrons scattered within the gun. Those leaving the muzzle continue with unchanged velocity because the wall of the gun, the target, the shield, and the outer wall of the collector are all maintained at the same potential and constitute another field-free inclosure. The electrons thus strike the target with the velocity determined by equation (9′). Because of the one-to-one correspondence existing between velocity and voltage, it is customary to express the velocity simply as so many "volts." However, a 200-volt electron does not have twice the velocity of a 100-volt electron, but, because of the radical, has only $\sqrt{2}$ times this last velocity.

By inserting in the De Broglie relation, equation (2), the numerical values of h and m, and the value of v given by equation (9′), it is possible to express the electron wave-length directly in terms of V, the bombarding potential:

$$\lambda = \frac{h}{mv} = \frac{6.55\times10^{-27}}{(9.00\times10^{-28})(5.95\times10^7\times V^{\frac{1}{2}})} = \frac{12.25\times10^{-8}}{V^{\frac{1}{2}}}\ \text{cm}\,, \tag{10}$$

$$\lambda = \frac{12.25}{V^{\frac{1}{2}}}\ \text{Ångstroms}\,; \tag{10'}$$

in equation (10′), λ is in Ångstroms and V is in volts.*

A gun of improved construction is now used, the one above described being an early type. In the improved gun a plate P_1 (Fig. 231), with an opening in it, is placed abreast of the filament and charged somewhat more negatively than this. This plate repels the electrons and tends to prevent them from traveling laterally to the wall. Another plate, P_2, is charged to a high positive potential for the purpose of giving additional directive effect. The potential of the plate P_2 is more highly positive than the wall. The electrons are thus given a high initial velocity in order to direct them toward the

* For potentials of thousands of volts a relativity correction to eqs. (9) and (10) becomes of appreciable magnitude—as further explained in sec. 12.

apertures and are then slowed down again. This procedure causes *many more* to pass through the apertures, and there results a more intense beam. The *velocity* of the stream as it enters the aperture in the plate P_3 and from that point on is, however, unaffected by the potentials of the plates P_1 and P_2. The velocity of the stream which is incident upon the target is still determined by V, the difference of potential between the filament and the wall of the gun.

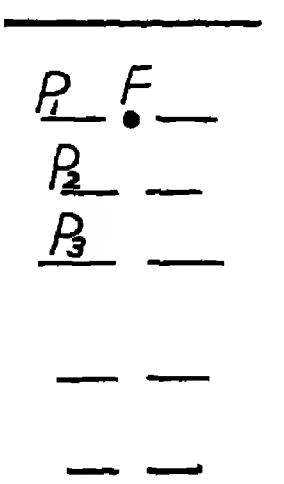


FIG. 231.—Improved electron gun.

The collector is a double Faraday box, the inner and outer boxes being insulated from one another by quartz plates. The lead wire from the inner box is especially shielded from stray currents by incasing it in quartz tubing. A potential is applied to the inner box which is negative in relation to that of the outer box. The field between the inner and outer boxes thus retards the electrons which enter the outer aperture. Those which have sufficient velocity to "make the grade" pass on into the inner box and contribute to the collector current. The remainder are turned back. Denote the potential difference between the collector boxes by V_c. Then all of those electrons which leave the target with velocities which, measured in volts, are greater than V_c will enter the inner box and those having lower velocities will be excluded.* The collector current is measured by a sensitive galvanometer. By making the potential difference at the collector just slightly less than that in the gun, it is possible to limit the collector current to electrons which have lost little or no speed or kinetic energy, in being deflected at the target. These so-called "full speed" electrons constitute an important group. When measurements upon them are being performed, the potential difference in the collector is always adjusted to be a little less, let us say 10 per cent less, than that in the gun. This allows for about 5 per cent range in velocity below the calculated value, which must be allowed for because of the effects mentioned in the previous footnote.

The full-speed electrons are the only ones for which diffraction effects have been demonstrated. Moreover, such effects are not expected for those electrons which have given up a portion of their energy toward raising atoms of the target to excited energy states. Electrons which have lost velocity are said to have made "unelastic collisions." These are, as a rule, excluded from the collector or other receptor. When not excluded they contribute only to the diffuse background, which is present in any case.

The entire apparatus of Davisson and Germer is small. The distance

* Because of edge effects at the apertures, the "cut off" is not absolutely sharp. Further, the incident beam is not strictly homogeneous as to velocity chiefly because of the slight fall of potential which occurs along the filament due to the current which heats this.

from the end of the gun or the collector to the target is of the order of a centimeter. The diameters of the various apertures are of the order of a millimeter. The difficulties of making measurements with a "spectrometer" having a length of arm of only a centimeter are very great.

The experiments of Davisson and Germer were carried out with single crystals of nickel.

4. The Nickel Crystal.—Nickel crystallizes in a face-centered cubic lattice (see Fig. 232, left). When a corner is cut from the cube at right angles to a body diagonal, or cube diagonal, as shown on the right, a (1 1 1) face is exposed. This is a face which occurs commonly and in relative perfection in

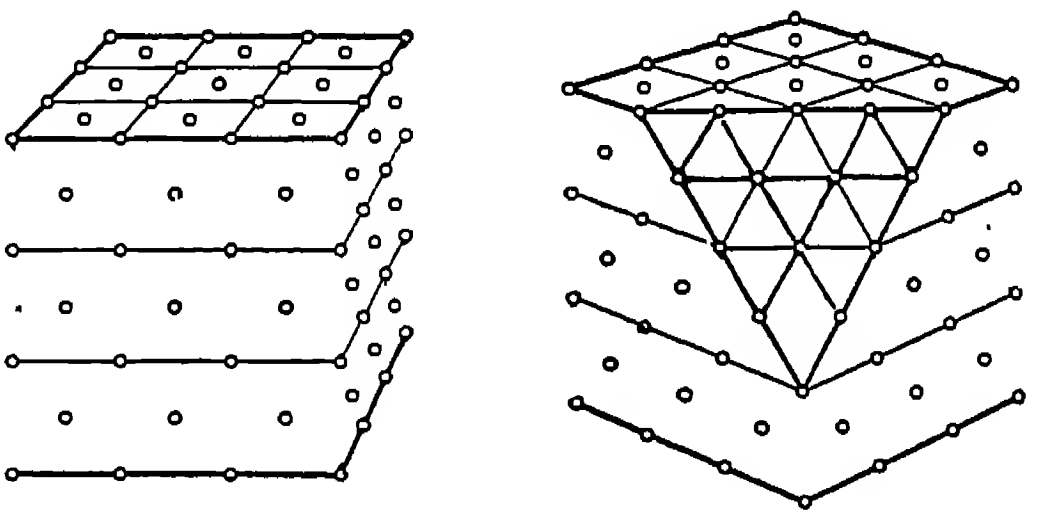


FIG. 232.—(From Davisson and Germer)

crystallized nickel. It is such a face which was used by Davisson and Germer. In it the atoms are arranged in equilateral triangles. The face is bombarded perpendicularly, as shown in Fig. 233. The angle θ between the collector and gun is called the "colatitude," this being the complement of the "latitude"—the angle between the collector and the crystal face. Provision is made for rotating the crystal face in its own plane, about a vertical axis in the figure. This is rotation in "azimuth." The crystal is shown in an "*A*-azimuth" with reference to the plane determined by the axis of the gun and the axis of the collector.

The upper part of Figure 234 is a view looking down on a (1 1 1) face of the cube. The atoms of the surface layer, those of the (1 1 1) face, are represented by the circles numbered *1*. The atoms of the second layer occupy the positions numbered *2*. Those on the cube faces are indicated by heavy numerals. Those of the second layer in the interior of the crystal (light numerals) fall under the centers of alternate triangles formed by the atoms of the first layer. A model of the lattice will be found helpful toward comprehending these facts and should be obtained if feasible, or a model may perhaps be improvised by the reader. Continuing, atoms of the third layer occupy positions numbered *3*. Those indicated by light numerals fall under the centers of the remaining alternate triangles formed by the first layer. Atoms

of the fourth layer fall directly under those of the first layer, and atoms of the fifth layer fall under those of the second, etc.

The three azimuths through the vertices of the triangular face are termed "A-azimuths"; one of these is lettered A in the figure. The three opposite azimuths, those through the edges of the face, are termed B-azimuths; one of these is lettered B. The lower portion of the figure shows the *layers*

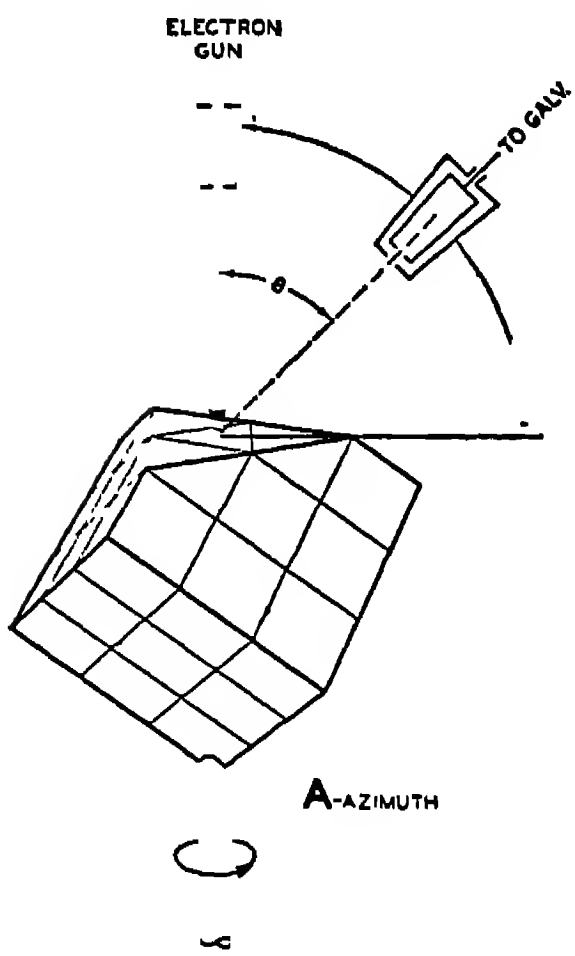


FIG. 233.—Perpendicular bombardment of (1 1 1) face (Davisson and Germer).

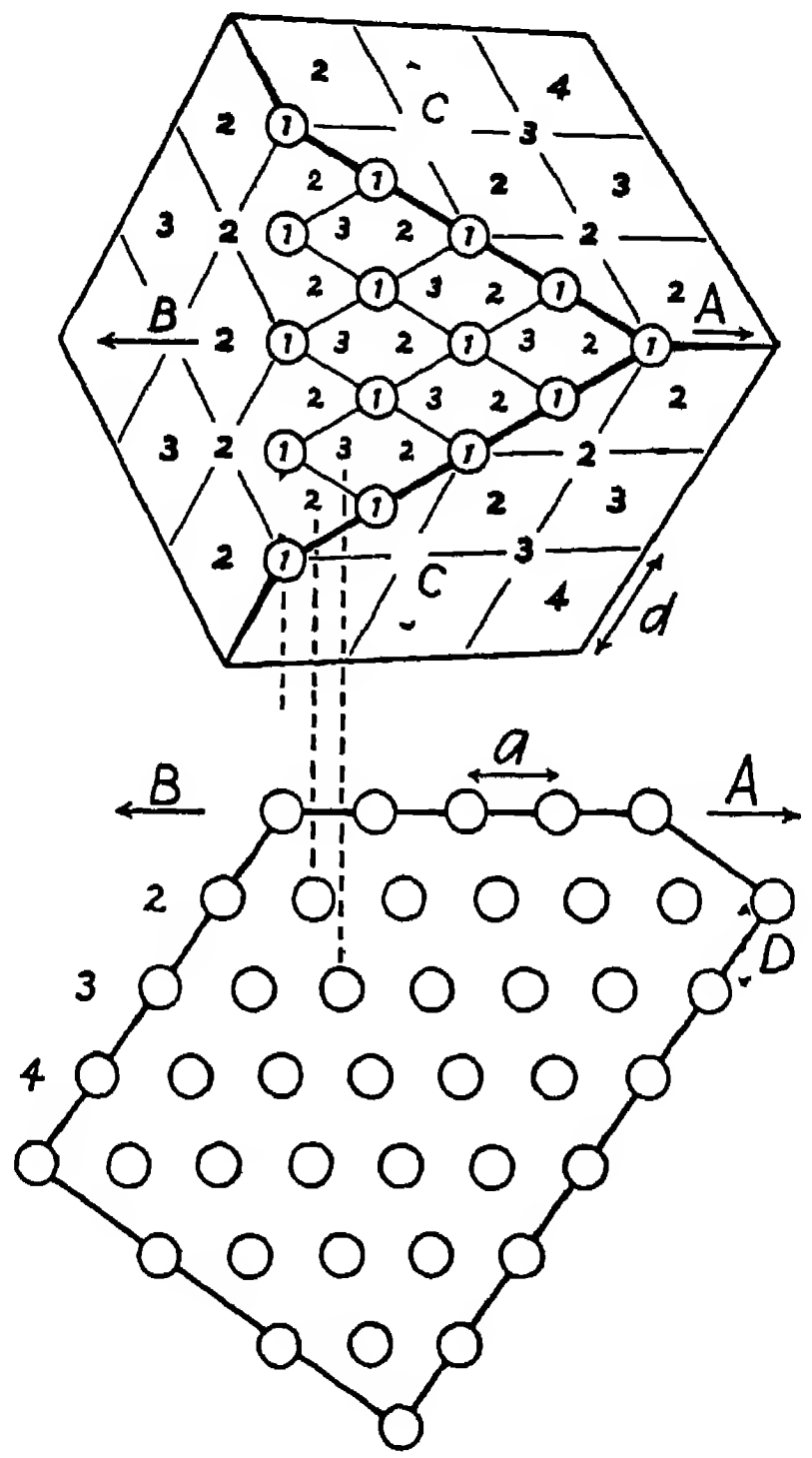


FIG. 234

of atoms, the rows of atoms being viewed end on. This refers to the rows perpendicular to an A- or B-azimuth. The rows of the second layer are displaced relative to those of the first layer by one-third the distance a, the distance between rows. This displacement is in the direction of the A-azimuth, or to the right in the figure. The rows of the third layer are again displaced to the right by the distance $a/3$, etc. It is to be noted incidentally that every lattice point has neighbors in the same relative position as has every other lattice point. The three A-azimuths differ physically from the opposite azimuths, the B-azimuths, because of the manner in which the

rows of atoms in the successive layers are displaced—by $a/3$ in the A direction or by $2a/3$ in the B direction. The two azimuths marked C, above, are parallel to an edge of the face. The atoms form rows perpendicular to these azimuths also. When these rows are viewed end on, the rows in the various layers fall directly under each other. Hence opposite azimuths are physically indistinguishable. There are consequently six C-azimuths, an opposed pair parallel to each edge of the triangular face. The A-, B-, and C-azimuths, twelve in all, are called the "principal azimuths." The diffracted beams occur in the principal azimuths. Alternative designations for them, based on Miller indices, are: (1 1 1) for the A-azimuths; (1 0 0) for the B-azimuths; and (1 1 0) for the C-azimuths. The manner in which these designations are arrived at does not concern us.

TABLE VIII

Length of edge of unit cube	$d=3.510$ Å
Separation of (1 1 1) planes	$D=\frac{d}{\sqrt{3}}=2.03$
Length of side of triangles in (1 1 1) planes	$\frac{d}{\sqrt{2}}=2.48$
Distance between rows:	
In A- or B-azimuths	$a=\frac{d}{2}\sqrt{\frac{3}{2}}=2.15$
In C-azimuths	$c=\frac{d}{2\sqrt{2}}=1.24$

The dimensions of the nickel crystal lattice are as follows: The length of the edge of the unit cube is $d=3.510$ Å. The length of the body diagonal of the unit cube is $\sqrt{3}d$. The separation between layers—that is, between (1 1 1) planes— is one-third of the length of the body diagonal, or $D=\frac{1}{3}\sqrt{3}d=2.03$ Å. The length of side of the equilateral triangles formed by the atoms in a (1 1 1) plane is $d/\sqrt{2}=2.48$ Å. The distance between rows of atoms in the A- or B-azimuths, denoted by a, is $a=\sqrt{(2.48)^2-(2.48/2)^2}=(2.48/2)\sqrt{3}=2.15$ Å. The distance between the rows of atoms in the C-azimuths will be denoted by c and $c=2.48/2=1.24$ Å. These data are summarized in Table VIII.

5. Voltage, Colatitude, and Azimuth Curves.—In discussing measurements we shall suppose that the current due to the incident stream of electrons is constant. The number of full-speed electrons recorded by the collector is then a function of three, and only three, independent variables: first, the voltage of the incident beam or, in other words, the velocity or wave-length; second, the colatitude of the collector; and, third, the azimuth setting. Exploration is made by varying one of these factors at a time.

There result then respectively "voltage curves," "colatitude curves"—for example, Fig. 229—and "azimuth curves." In each type of curve the collector current, which measures the intensity of the beam, is plotted as dependent variable, that is, as the ordinate.

The broken curve in Figure 229 represents background scattering which is always present. This has a steep slope. As a result, some beams, which are not among the strongest, are masked, and fail to appear as well-marked lobes or spurs in colatitude curves. When voltage curves are taken, the existence of these beams may become manifest.

6. Plane-grating Beams.—In Figure 235 several voltage curves are shown. The crystal is throughout in a C-azimuth. The collector is set at a given large colatitude angle (see upper portion of figure), for example, at $\theta = 80°$. The collector current is determined at each of a series of bombarding potentials. As the bombarding potential is raised the beam sweeps across the collector opening, as indicated by the arrow. A maximum occurs in the voltage curve at 100 volts. According to the De Broglie theory, the associated wave-length for 100-volt electrons is, by equation (10′), $12.25/(100)^{\frac{1}{2}} = 1.225$ Å. Does this agree with the wave-length which we would expect for waves diffracted by the surface layer of nickel atoms considered as a plane grating? To determine this, consider the rows of atoms which are perpendicular to the C-azimuth as constituting the lines of a grating. The wave-length to be expected is $\lambda = c \sin 80° = 1.24 \times .985 = 1.22$ Å, in agreement with the De Broglie theory.

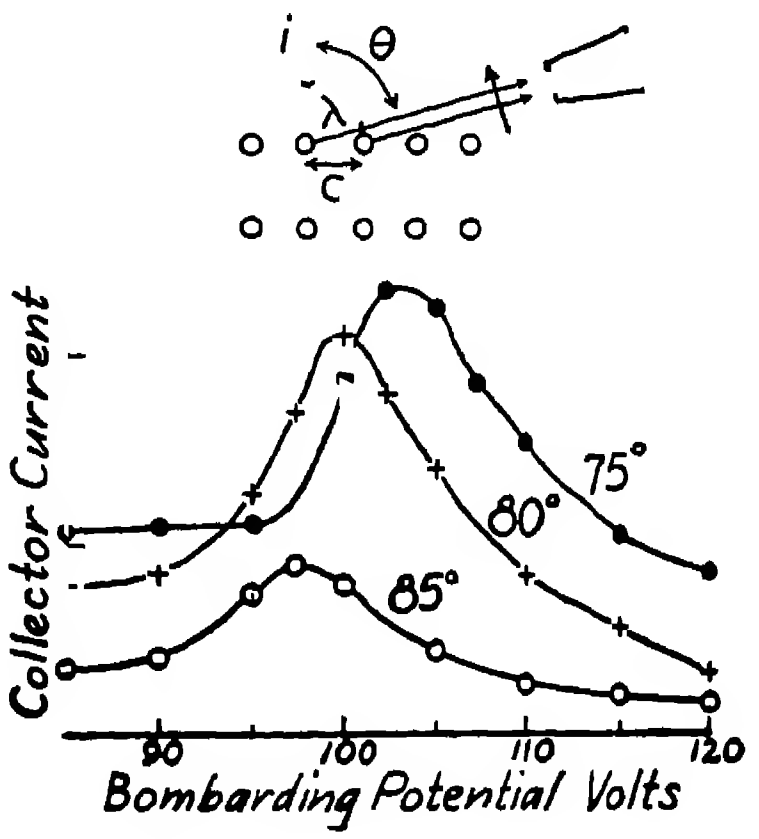


FIG. 235.—Below, voltage curves for plane-grating beams in C-azimuth (Davisson and Germer). Above, rows of atoms considered as a plane grating.

For a colatitude setting of 75° the maximum in the voltage curve shifts to the right—the voltage required to produce the beam is greater, the wave-length is smaller. The shift accords in direction and magnitude with the De Broglie theory and the plane-grating formula as the reader may determine for himself by taking data from the curve. For colatitudes below 75° the maxima in the curves exhibit a progressive decrease in intensity, and plane-grating beams are not observed at all below 60° colatitude. The decrease in intensity presumably arises from destructive interference caused by energy scattered by layers of atoms below the surface layer.

At 85° colatitude the maximum in the curve occurs at 97.5 volts. The condition is nearly that of grazing emergence because sin 85° = .996, or nearly 1. When radiation is incident upon a grating perpendicularly, the greatest wave-length which can be diffracted is $\lambda = c$, the grating space. The diffracted beam then emerges at grazing angle. This means that no beams should be found in the C-azimuths below 97 volts, and they are in fact not found.

Plane-grating beams also occur in the A- and B-azimuths. The equivalent plane grating is the same for these two sets of azimuths, which differ only in regard to the relative position of the lower layers of atoms. The beams appear at grazing emergence at 32.5 volts, which is the voltage at which they would be expected. As the voltage is raised, the beams shift toward smaller colatitudes, obeying the plane-grating relation.

Beams near grazing emergence are observed only when the vacuum is very high and the surface of the crystal has been freed from adsorbed gas by heating. This refers to plane-grating beams arising from diffraction by atoms of the crystal; for others see the next section.

7. Gas Beams.—When a small amount of gas is present upon the crystal surface, other plane-grating beams appear. These originate from a grating which has double the expected spacing, that is, double the spacing to be expected in a given azimuth from the surface layer of nickel atoms. The presumption is that a "net" of gas atoms or molecules is formed which has twice the scale factor of the nickel surface-lattice. A gas atom located over the center of every fourth triangle of the nickel lattice would satisfy this requirement (see Fig. 236). The gas atoms, or molecules, are believed to be so located.

As gas continues to collect, the gas beams of the plane-grating type fade out. There then arise beams due to numerous layers of gas atoms. These have also been studied.*

8. Space-Lattice Beams.—Several strong beams (of which the one shown in Fig. 229 is an example) occur in each of the principal azimuths. Each beam appears in full development when, and only when, the *speed of bombardment has a definite value which is characteristic for the beam.* This fact is illustrated by the series of curves shown in Figure 237. Each of these curves is taken for a different speed of bombardment, that is, at a different voltage. The beam represented in these curves attains its full development at a bombarding potential of 54 volts, and it is hence known as a "54-volt beam."

A given speed of bombardment represents a given wave-length. Thus beams occur for discrete wave-lengths in discrete directions. This fact pre-

* See Davisson and Germer, *Phys. Rev.*, **30**, 705, 1927; L. H. Germer, *Zeit. f. Phys.*, **54**, 408, 1929, or translation in *Bell Syst. Tech. Jour.*, **8**, 591, 1929; E. Rupp, *Ann. d. Phys.*, **5**, 453, 1930; and H. E. Farnsworth, *Phys. Rev.*, **40**, 684, 1932.

sents a similarity to X-ray diffraction and establishes a presumption that we have here to do with diffraction by several layers of atoms, that is, with diffraction by a "space lattice." This presumption will receive more and more support as we proceed. The foregoing 54-volt beam is thus to be re-

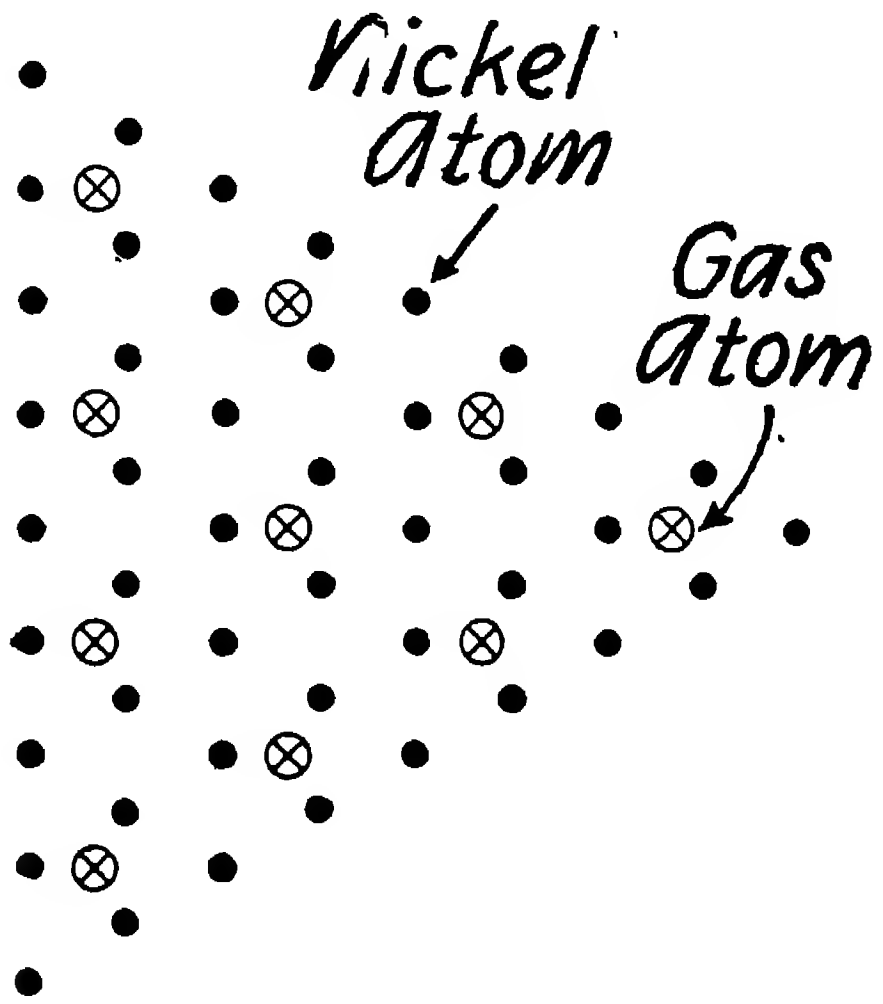


FIG. 236.—Layer of adsorbed gas atoms (Davisson and Germer)

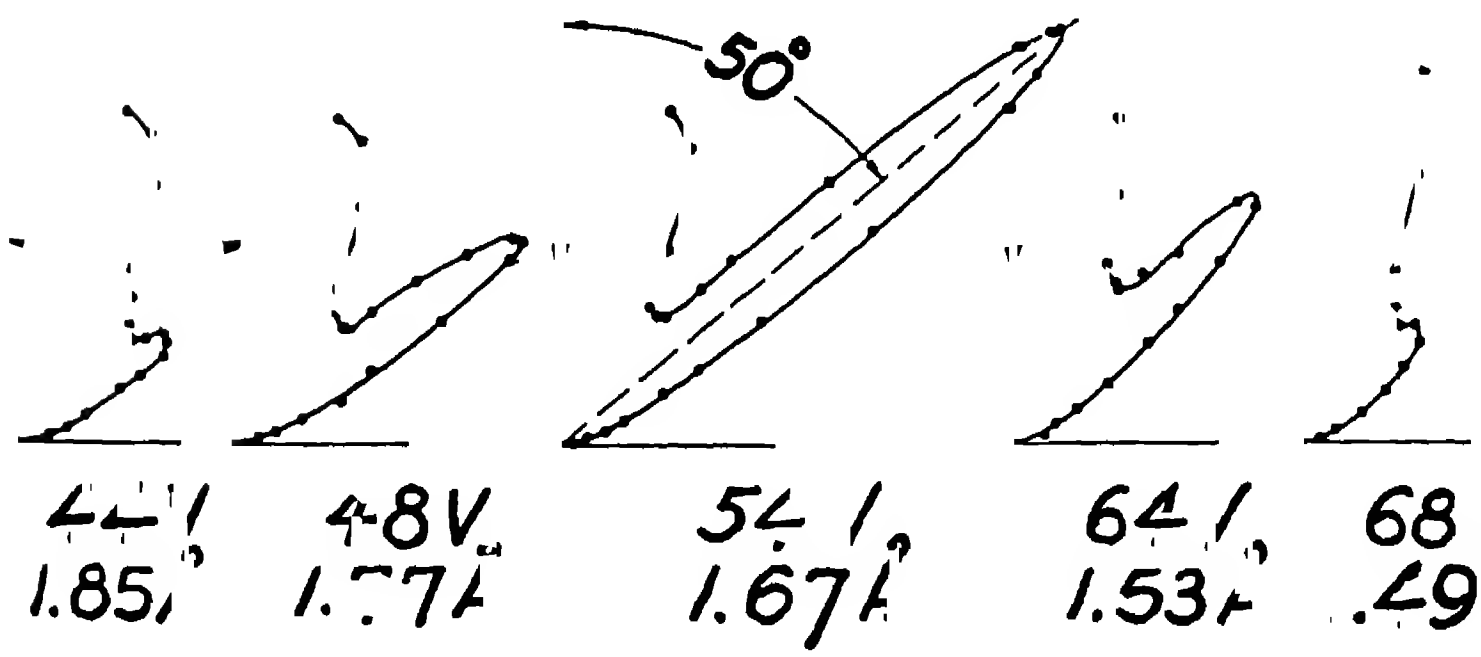


FIG. 237.—Growth and decay of 54-volt beam (Davisson and Germer)

garded as the electron-wave analogue of a corresponding Laue X-ray beam. An important fact is to be borne in mind, however. In the Laue experiment "white radiation" is incident, and, consequently, there may be numerous quite distinct beams issuing from the crystal at once. In electron diffraction experiments on the other hand, the incident beam is homogeneous, or approximately so—the incident wave train must hence be regarded as ap-

proximately monochromatic—and in accordance with this fact, only when the bombarding potential has one of certain definite values will there be a beam in full development, and then there will be only a single beam.

Electrons of speeds up to 370 volts were used in the experiments of Davisson and Germer which we are discussing. These are classed as slow electrons. Their power of penetrating matter is known to be very low. Consequently it is to be expected that the number of atom layers by which they are diffracted should prove to be small. It is perhaps only three or four. This in turn means low "resolving power." The resolving power depends upon the number of atom layers in somewhat the same manner as it depends upon the number of ruled lines for a diffraction grating. And as a result of the low resolving power, the wave-length of the incident beam, or, in other words, the voltage, is not as critically determined as it would be if the resolving power were higher. The 54-volt beam has appreciable intensity over a voltage range from 44 to 68 volts, as shown by the various curves of Figure 237. This corresponds to a wave-length range of 11 per cent in either direction from the maximum.

The 54-volt beam moves to smaller colatitudes as the voltage rises, as would be expected. The same is not true, however, of all space-lattice beams, for reasons not fully known.

Figure 238 shows examples of space-lattice beams as revealed when azimuth curves are taken. The lower curve was taken at 50° colatitude and 54 volts bombarding potential. Under these conditions the 54-volt beam in the A-azimuth is fully developed, and hence this beam shows as an intense maximum in each of the A-azimuths. The less intense maxima in the B-azimuths are remnants of another beam (see upper curve). The upper curve was taken at 44° and 65 volts, under conditions of full development of a "65-volt beam" which occurs in the B-azimuths, and this beam here shows as an intense maximum in each B-azimuth. The less intense maxima in the A-azimuths are remnants of the 54-volt beam, which still shows as an appreciable spur at 65 volts, and a portion of this spur is admitted into the collector opening at 44° colatitude.

The experiments thus far described are those of Davisson and Germer on diffraction from a (1 1 1) face of a nickel crystal.*

H. E. Farnsworth† has carried out similar experiments upon (1 0 0) faces of copper, silver, and gold crystals. He has improved the technique, obtaining sharper, narrower beams, and has obtained numerous additional interesting results. Some of these present extremely perplexing problems.

* *Phys. Rev.*, **30**, 705, 1927.

† *Ibid.*, **34**, 679, 1929; **35**, 1131, 1930; **40**, 684 and 1049, 1932; **42**, 588, 1932; and **43**, 900, 1933.

W. T. Sproul* has bombarded the (1 1 2) and (1 0 0) faces of a tungsten crystal. He has modified the apparatus, introducing the magnetic deflection method in an ingenious manner.

9. The Refraction of Electron Waves.—*X-rays* travel with almost the same velocity within matter as they do in vacuum, and there is consequently

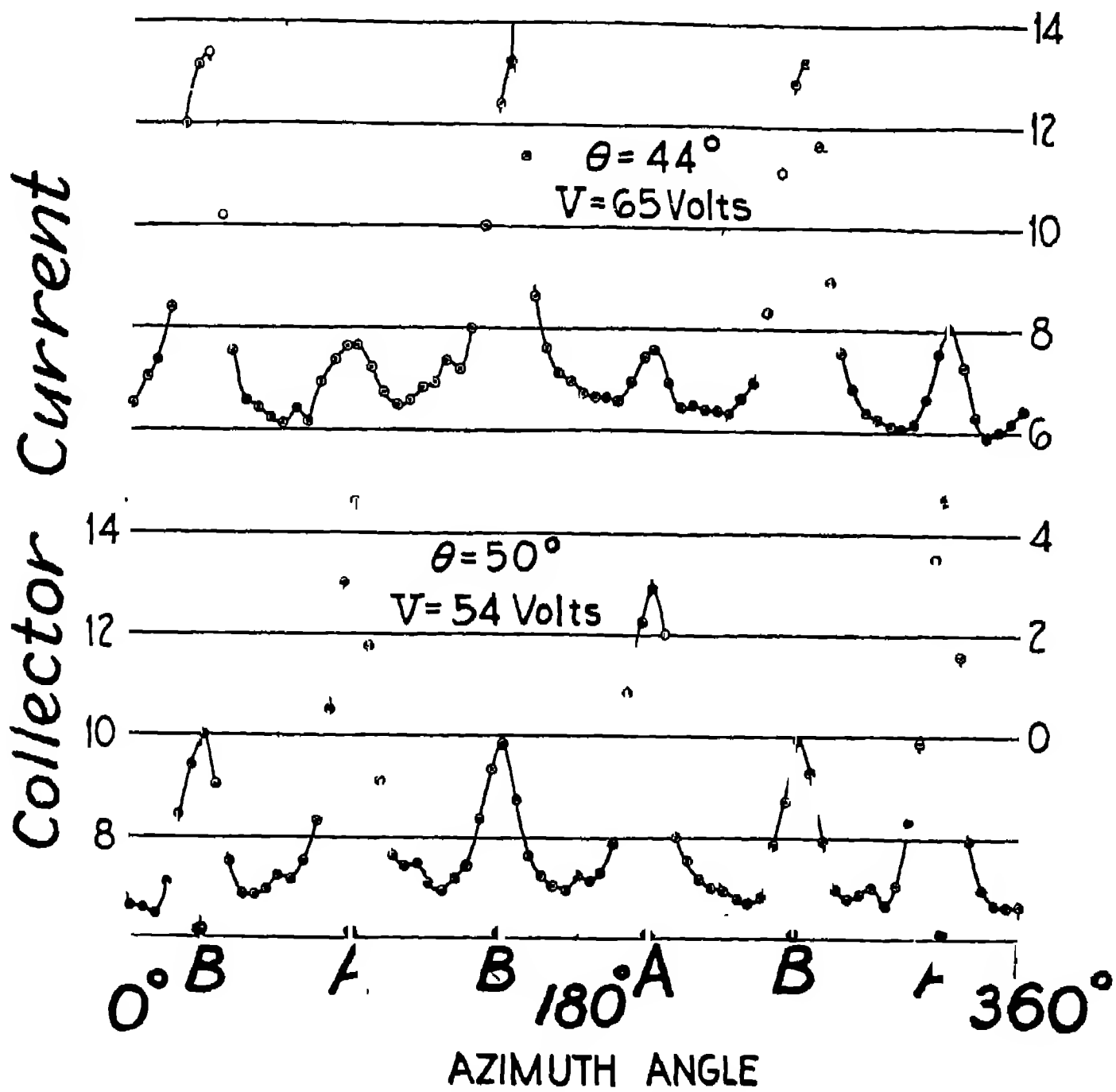


FIG. 238.—Azimuth curves (Davisson and Germer)

no refractive effect, or only an extremely minute one. For electron waves the situation is different, especially when the electrons are moving slowly. However, the knowledge which we have regarding the refractive effect of matter upon electron waves is rather fragmentary. Such as it is, it has been furnished by the very electron-diffraction experiments which we are discussing:

Referring to Figure 239, let A_x, B_x, C_x, D_x, and E_x represent the Laue *X-ray* beams which would be diffracted backward by a certain crystal. If the crystal exerted no refractive effect upon *electron-waves* the corresponding electron beams would attain their full development at the same wave-

* *RSI*, **4**, 193, 1933, and *Phys. Rev.*, **43**, 516, 1933. See also discussion of the results by H. E. Farnsworth and by G. P. Thomson, *ibid.*, **44**, 417, 1933.

lengths, as computed from the speed of bombardment by the De Broglie relation, equation (10), and would take these same directions. Instead, the electron beams, A_e, B_e, etc., attain their full development at a somewhat lower speed of bombardment—that is, at greater wave-length—and are diffracted to correspondingly larger angles, as indicated.

Since we are concerned with normal incidence, the plane-grating law takes the form $m\lambda = a \sin \theta$, where a is the (horizontal) distance between the atom rows which are perpendicular to the plane of the figure. Evidently the X-ray

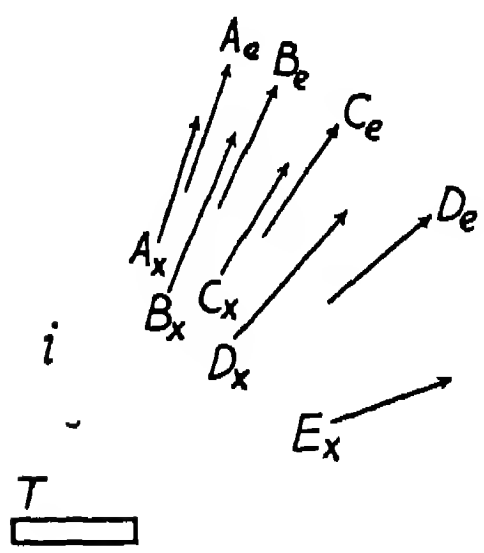


FIG. 239.—Laue X-ray beams A_x, B_x, etc., and electron beams, A_e, B_e, etc.

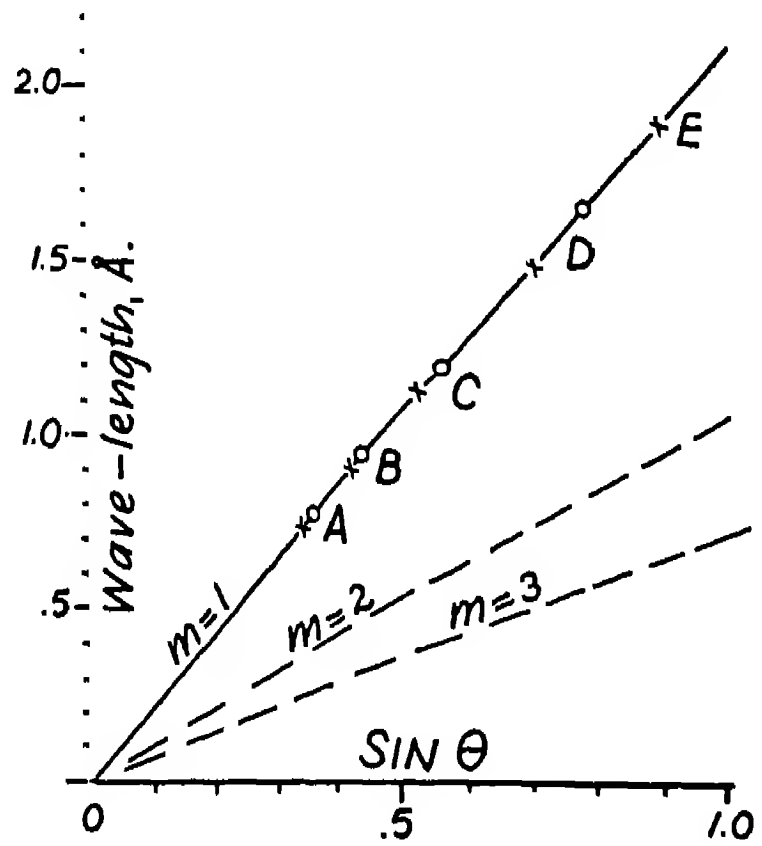


FIG. 240 —The plane-grating relation. X-ray beams, crosses; and electron beams, circles.

beams, while they are space-lattice beams, must obey the plane-grating law. Hence, taking m as unity, and plotting $\sin \theta$ as abscissa (see Fig. 240) and λ as ordinate, the X-ray beams must fall on that straight line through the origin, which has for its slope $\lambda/\sin \theta = a$. They fall in particular at the points marked by crosses at A, B, C, etc., when we are dealing with a nickel crystal in an A-azimuth. That they must occupy these positions may be shown from the geometry of the nickel crystal, as carried out in detail in Appendix H, where several matters with which we are concerned in the present section and the next one are more fully discussed.

Now it is found that when, for each of the electron beams, $\sin \theta$ is plotted as abscissa and λ as calculated from the De Broglie relation is plotted as ordinate, the electron beams fall upon the same straight line with the X-ray beams. The positions of the electron beams in this graph are indicated by circles at A, B, etc. Thus the electron beams satisfy the plane-grating law, and the fact that they do satisfy this law constitutes a verification of the

De Broglie relation for the space-lattice beams even though these beams are refracted, that is, have a wave-length within the crystal which differs from that in vacuum. This appears more fully in Appendix H.

Be it stated incidentally that there is no electron beam emerging from the crystal which corresponds to the X-ray beam E_x. This is because this electron beam, within the crystal, strikes the surface of the crystal at such a large angle of incidence that it suffers "total internal reflection."

The above-considered refraction of electron waves may be accounted for in a general way, and the amount of this refraction may be calculated. The agreement between observation and calculation is partially very good, but there is also much lack of agreement. Because the agreement is incomplete, the theory of the refraction, which will now be briefly given, must be regarded as representing only a first, rough approximation:

Each atom of the crystal which is under bombardment consists of a positive nucleus and several surrounding electron shells. Let the crystal, as a whole, be at zero potential. At a point within the outer electron shell of a given atom a small positive potential will nevertheless prevail, and at a point within the first two electron shells a somewhat higher positive potential will prevail. Thus, although the potential of the crystal considered in the gross is zero, there must nevertheless be a certain potential, in general different from zero, ascribed to each point within the crystal; this potential is known as the "inner potential" or "grating potential" at that point. Because of the existence of this potential, an electron incident upon a crystal will be accelerated in penetrating the electron shells of the atoms, and will on the average travel faster within the crystal than in vacuum. Now, following H. Bethe,* the "average grating potential," which we shall designate by V_g, may be taken as a basis for calculating the gain in velocity experienced by an electron considered as a wave packet.

It is possible to evaluate V_g for, for example, nickel, from the known mean kinetic energy possessed by a free electron within the metal and the known work which is necessary to remove an electron from the metal. Upon this basis V_g for nickel comes out to be about 17 volts.

The total fall of potential which an electron has experienced after it has penetrated into the crystal is $V' = V + V_g$, where V is the bombarding potential. If v is the velocity of the electron in vacuum, applying equation (9), the velocity within the crystal is:

$$v' = v\sqrt{\frac{V+V_g}{V}}. \tag{11}$$

* *Naturwiss.*, **16**, 333, 1928. Later related articles by the same author are to be found in *Annalen* (4), **87**, 55, 1928, and (5), **5**, 325, 1930. In the same connection also see C. Eckart, *Proc. Nat. Acad.*, **12**, 640, 1927, and H. Bethe, *Naturwiss.*, **15**, 787, 1927.

The wave-length associated with the electron within the crystal is then, according to equation (10), equal to:

$$\lambda' = \frac{h}{mv'} = \frac{h}{mv}\sqrt{\frac{V}{V+V_g}} = \lambda\sqrt{\frac{V}{V+V_g}}\,. \tag{12}$$

That is, owing to the higher velocity of the electron within the crystal, the wave-length is *less* than in vacuum. (Correspondingly, according to eq. [5], the velocity of the phase waves is *lower*.)

We are now in a position to comprehend more thoroughly why the space-lattice electron beams attain full development at a greater wave-length, and are consequently diffracted to greater angles, than the corresponding X-ray beams. The electron beams attain full development when the wave-length in vacuum (speed of bombardment) is such that after the electron has been accelerated upon entering the crystal, the then shorter, associated wave-length is the same as the wave-length of the corresponding X-ray beam. Within the crystal the directions of the diffracted X-ray and electron beams are the same—and upon leaving the crystal the electron beams are deviated away from the normal.

As in ordinary optics, the wave-length in vacuum divided by the wave-length in the refracting medium is equal to the sine of the external angle divided by the sine of the internal angle, and, in analogy with ordinary optics, this ratio for electron waves is defined as the *refractive index* of the medium for electron waves. Denoting this index by μ and making use of equation (12), we have:

$$\mu = \frac{\lambda}{\lambda'} = \frac{\sin\theta}{\sin\theta'} = \sqrt{\frac{V+V_g}{V}}\,. \tag{13}$$

The refractive index, μ, plotted as ordinate, against the wave-length λ, plotted as abscissa, will lead to a so-called "dispersion curve," and the relations expressed by equation (13) may be used as a basis for plotting an "observed" dispersion curve and a "calculated" dispersion curve, and comparing these. The calculated curve is obtained by plotting the function $\sqrt{(V+V_g)/V}$ as ordinate, using V_g as a constant, namely, 17 volts. The observed curve might conceivably be obtained by plotting as ordinate $\sin\theta/\sin\theta'$, where θ is the observed angle of diffraction at full development of the beam, and θ' is the corresponding angle of diffraction for an X-ray beam. But because of the impossibility of measuring angles accurately with a "spectrometer" which has an arm as short as has the "electron spectrometer," this method would not be very accurate. An alternative method is to plot as ordinate the values of λ/λ', where λ is the wave-length at which

the beam is fully developed, and λ' is the corresponding X-ray wave-length; the values of λ for this purpose can be best obtained by using the method of Bragg reflection, and Davisson and Germer carried out a separate research by this method for the purpose of obtaining the desired data.

10. The Bragg Reflection of Electron Waves.—In applying the method of Bragg reflection, or diffraction, the angle of incidence and the angle of reflection are made equal and are held fixed, and the crystal remains fixed. The bombarding potential alone is varied.*

The Bragg view considers space-lattice diffraction as constructively interfering reflection from successive atom layers. This view may be applied to electron waves. However, because of the refraction of these waves, Bragg's law in its ordinary, simple form does not apply; this law in its simple form assumes that $\mu = 1$. Accordingly, it becomes necessary to derive a more general expression which will hold even when μ differs from unity to a marked degree. This is carried out in Appendix H for the benefit of those who may be interested in following the derivation. Suffice it here to direct attention to the fact that this expression must perforce contain both λ and μ as unknowns. However, if we assume that the De Broglie relation has already been sufficiently established, as we well may, we may consider λ to be given by the bombarding potential, and then deduce the values of μ from the results of experiment.

In the experiments in question the bombarding potential was varied by steps and the collector current was measured at each step. The curve shown in Figure 241 was obtained. Here the square root of the bombarding poten-

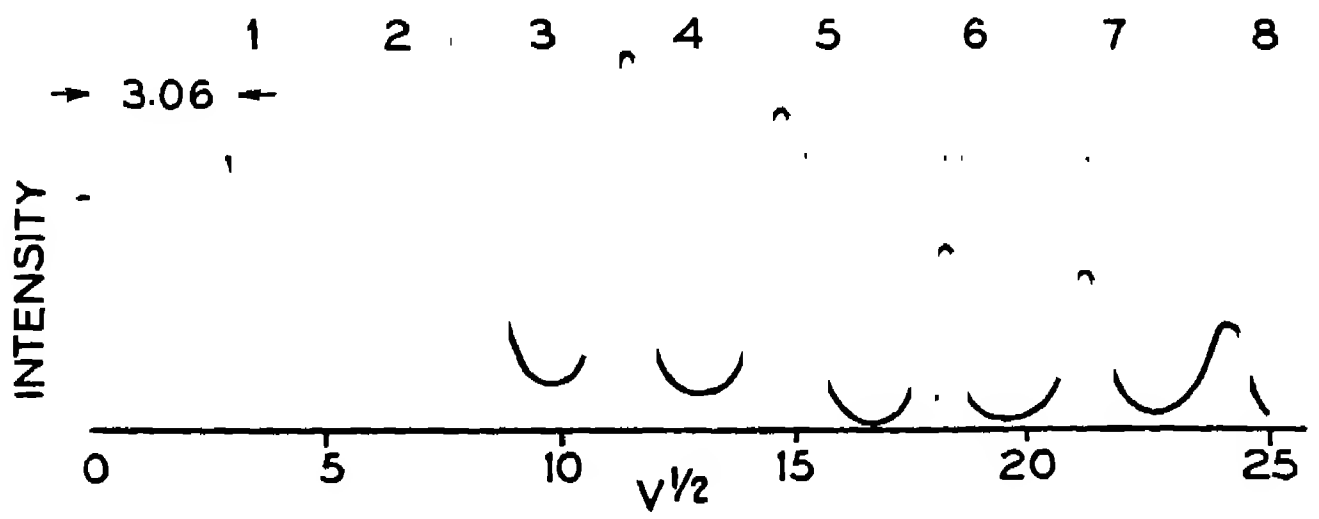


FIG. 241.—Voltage curve for Bragg reflection (Davisson and Germer)

tial is plotted as abscissa and the intensity, collector current, is plotted as ordinate. This curve shows that Bragg reflection, which is really space-lattice diffraction, occurs, as is to be expected, at only definite bombarding potentials, as indicated by the maxima of the curve. Moreover, from the positions of the actual maxima and the voltages at which the beams would de-

* See Davisson and Germer, *Proc. Nat. Acad.*, **14**, 317 and 619, 1928.

velop if there were no refraction, which are indicated in the figure by the downward-pointing arrows, a series of values of $\mu = \lambda/\lambda'$ may be deduced. (The number to the right of each arrow gives the value of the spectral order, m. The details for calculating the voltages at which the beams would develop if there were no refraction are given in the last paragraph of Appendix H.)

When the observed values of μ, obtained as just explained, were plotted as a dispersion curve, the resulting curve was found to be in good agreement with the curve calculated from the inner potential, over a good portion of the explored range, but failed to agree over another portion.

The wave-length region in which there is lack of agreement has sometimes been interpreted as representing a region in which there occurs something akin to "anomalous dispersion."

H. E. Farnsworth* has in his careful measurements found lack of agreement also for copper, silver, and gold, and he expresses this by saying that the inner potential of the crystal is not constant but is smaller for the lower velocities.

The problem has also been treated theoretically. It develops that the inner potential cannot strictly be considered as a constant of the material.†

Much work will be required to clear up the whole question.

Bragg reflection experiments have also been carried out with crystals which are insulators, such as rocksalt, serving as targets, but difficulties arise in connection with dissipating the static charge which accumulates due to the electron bombardment. The results obtained with low-speed bombardment are not reliable.‡

By inspection of equation (13) it becomes evident that the refractive index for electron waves approaches unity as the bombarding potential is raised. Consequently, refractive effects will be less marked for high-speed electrons, and may under suitable conditions be negligible. Nevertheless, investigations upon the refraction of electron waves may to advantage be carried out with high-speed streams, because in experiments with such streams photographic methods may be employed with a consequent increase in the accuracy of measurement. Moreover, high-speed streams have the advantage of being less subject to annoying disturbing effects of various types.§

* *Phys. Rev.*, **40**, 684, 1932.

† P. M. Morse, *ibid.*, **35**, 1310, 1930; M. von Laue, *Sitzungsber. d. Ber. Akad.* (1930), p. 26.

‡ See, e.g., Rupp, *Ann. der Phys.*, **3**, 497, 1929; von Laue and Rupp, *ibid.*, **4**, 1097, 1930; and, for discussion, F. Kirchner, *Ergebnisse d. Ex. Naturwiss.*, **11**, 81, 1932.

§ See, e.g., G. P. Thomson, *Proc. Roy. Soc.*, **133**, 20, 1931, and H. Raether, *Zeit. f. Phys.*, **78**, 527, 1932.

11. The Diffraction of Electrons by a Ruled Grating.—E. Rupp* has bombarded an optical grating with homogeneous streams of slow electrons, 70 and 150 volts. The grating was ruled on speculum metal and had 1300 lines per centimeter. The incident beam grazed the grating—the glancing angle was 5 or 10 minutes of arc. Beams near grazing emergence were received upon a photographic plate placed at some 40 cm from the grating. The central image beam was found to be intense. Close to it were the first-, second-, and even third-order diffracted images. These were faint but clearly discernible. The various images were separated by distances of about a millimeter. It is necessary to work at small glancing angles because it is only under these conditions that a ruled grating yields dispersion which is sufficiently large when such short wave-lengths are involved. The wave-lengths measured by Rupp confirm the De Broglie relation to within the errors of observation, 2 or 3 per cent.

B. L. Worsnop† has performed similar experiments.

THE DIFFRACTION OF FAST ELECTRONS

(Secs. 12–14)

12. Thomson's Method.—Electrons may also be diffracted by a method which basically resembles that of the powder spectrograph for X-rays, described in chapter 8, section 11. A homogeneous beam i (Fig. 242), issuing from the tube B, strikes an extremely thin film, F, commonly of metal, and passes through the film. Minute crystals in the film are oriented more or less in all possible directions. In this practically random orientation the film resembles the crystal powder used with X-rays. For good results the film must be so thin that each electron is scattered, diffracted, only once. A photographic plate is erected at P some 30 cm behind the film. Upon this plate appear circular rings which center about the point where the undeviated beam strikes the plate. The rings range in diameter from 1 to 3 cm.

The pioneer investigations by this method were made by G. P. Thomson, who began his work in 1926 and first announced results, with A. Reid, in June, 1927.‡

In Thomson's apparatus the electron stream originates at the cathode C of the tube A, which is operated by an induction coil. The tube A contains a small regulated amount of gas. It is necessary to have this gas present because a cold cathode is used; consequently, bombardment of the cathode by

* *Naturwiss.* (1928), p. 656; *Zeit. f. Phys.*, **52**, 8, 1928.

† *Nature*, **123**, 164, 1929.

‡ *Nature*, **119**, 890, 1927. Other papers of Thomson's giving accounts of his work by this same method are *Proc. Roy. Soc.*, **117**, 600, 1928; **119**, 651, 1928; **125**, 352, 1929; and *Phil. Mag.*, **6**, 939, 1928.

positive ions is required in order to liberate electrons from it. A variation of the method is to use a hot-wire cathode, as in modern X-ray tubes or in the electron gun of Davisson and Germer, and maintain a high vacuum in the chamber A. In either case, a portion of the electron stream passes through the fine tube B, of bore .23 mm and length 6 cm in Thomson's apparatus, and strikes the film F. The tube B is magnetically shielded by an iron tube not shown. The tube E is maintained at a high vacuum. Voltages ranging from 15,000 to 60,000 were used by Thomson. He found that the observed radii of the rings agree with radii computed by applying Bragg's law in its unmodified form. The voltages employed are so high that the refractive index of the film is nearly equal to unity, and, moreover, it may be shown that the remaining refractive effect is eliminated when the beam passes *through* the film.*

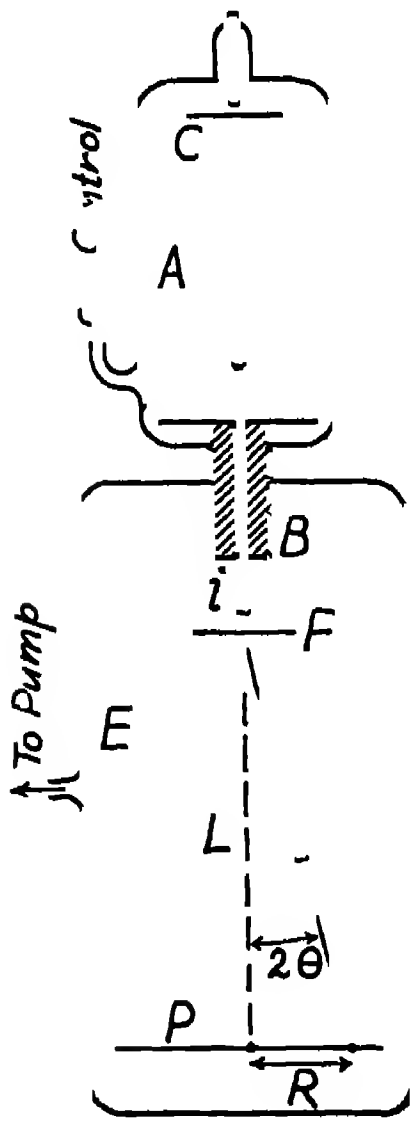


FIG. 242.—Thomson's method of electron diffraction.

When the voltage is varied, the radius of a given ring varies. As the voltage is raised the radius decreases. For a given ring reflection takes place in different microcrystals, but always in the same set of crystal planes, and therefore D, the spacing of the planes, is always the same for a given ring and so is m, the spectral order. The angle between the incident beam and the diffracted beam will be denoted by 2θ. Thus θ designates the glancing angle made by the incident and the diffracted beams with the reflecting net planes. The angles involved are all small. Applying Bragg's law and making approximations which are permissible for small angles:

$$m\lambda = 2D \sin\theta = D \tan 2\theta = \frac{DR}{L}. \tag{14}$$

Introducing the value of λ from equation (10′):

$$\begin{aligned} m\frac{12.25}{V^{\frac{1}{2}}} &= \frac{DR}{L}, \\ \frac{12.25\, mL}{D} &= RV^{\frac{1}{2}}. \end{aligned} \tag{15}$$

* See last paragraph of chap. 8, sec. 20, or, for detailed proof, G. P. Thomson, *Phil. Mag.*, **6**, 939, 1928.

As the voltage is varied the radius of a given ring varies in such a manner that the product of the radius and the square root of the voltage remains constant. For the faster electrons, in the upper portion of the voltage range, the relativity correction for the mass is made. The equation then obtained is:

$$\frac{12.25\, mL}{D} = RV^{\frac{1}{2}}\left(1+\frac{eV}{1200\, m_0c^2}\right). \tag{16}$$

The product of the radius and the corrected function of the voltage remains constant.* The correction never exceeds 3 per cent. Calculated and observed radii of the rings agree to about 1 per cent.

Displacement of the entire pattern, both direct beam and diffraction rings, may be brought about by applying a magnetic field perpendicularly between F and P. The occurrence of this displacement proves that the pattern is due to electrons and not due to X-rays generated somewhere by the cathode rays. The pattern shifts as a unit, without alteration or distortion. Hence, the various beams are each homogeneous, and the electrons in all of them have the same speed. This must, moreover, be "full speed," for there were numerous holes in the film used, thus assuring the presence of full-speed electrons in the central beam.

In Thomson's early experiments thin beaten metal foil was further thinned by immersion in a suitable reagent. The thickness was estimated at not more than 10^{-5} cm. The presence in the film of relatively large single crystals was sometimes revealed by the occurrence of especially intense spots along the diffraction rings. The rings were always rather broad. More recently films have been used which are formed by evaporation or cathode sputtering. They are deposited upon a base of rocksalt or cellulose which is later dissolved away. The metal film then drifts in the liquid and must be washed by transferring it in a spoon to a succession of baths. While in the bath it is maneuvered over a hole of about 2 mm diameter in a metal disk. Upon removal from the bath the film adheres to the disk. The free portion of the film, over the hole, is utilized, the disk serving as mounting. Films of the order of 10^{-6} cm are obtained. The thickness is thus only one-fiftieth of the wave-length of light. The films are highly transparent, sometimes almost invisible by transmitted light, but readily visible by reflected light. They are entirely too *thin* to show the common colors of thin films.

With the improved films the diffraction rings show as sharp "lines," the sharpness being comparable with that of X-ray spectrograms. Figure 243

* V is measured in volts; e is the electronic charge, 4.77×10^{-10} e.s.u.; m_0 is the rest mass of the electron, 9.00×10^{-28} gr.; c is the velocity of light, 3.00×10^{10} cm/sec.

shows a photograph obtained by Mark and Wierl* for silver. By using a cathode-ray tube especially designed for large current density, they obtain exposures in one-tenth and even one-fiftieth of a second.

Thomson's method, as thus far described, is limited to metal films.

F. Kirchner† has extended the method to a large variety of substances, for example, salts, by evaporating a minute quantity of the substance on to a very thin film of celluloid. The celluloid shows no outstanding pattern of its own. Patterns from films of celluloid, showing three rather broad, diffuse rings, were first obtained by A. Reid.‡

FIG. 243.—Pattern for silver obtained by Thomson's method (Mark and Wierl).

Thomson's method of electron diffraction as now developed is superior to the X-ray powder spectrograph as a means of investigating crystal structure under some circumstances—especially in investigating the structure of thin films.

Thomson, on the basis of his earlier results, calculated the apparent "resolving power" of his apparatus and concluded that "the fact that certain rings can be resolved shows that the electrons must be accompanied by a train of waves not less than 50 waves in length." On the basis of the superior resolution since attained by Mark and Wierl and numerous others, it is safe now to set this lower limit much higher, at perhaps 300 waves. According to modern views, the number of waves in the train is related to the degree of homogeneity which obtains in the incident stream of electrons. A highly homogeneous stream, corresponding to highly monochromatic light, would have many waves in a train.

13. The Diffraction of Electrons by Thin Crystals.—Soon after Thomson's first work, S. Kikuchi,§ under the guidance of S. Nishikawa, employed Thomson's method for the purpose of diffracting electrons by thin sheets of muscovite mica, $KH_2Al_3(SiO_4)_3$. He obtained several interesting patterns, of which one especially was for several years difficult to interpret. All are now, however, fairly well understood.

Provision was made in the apparatus for inclining the sheet of mica with reference to the incident beam. Various sheets were used, differing in thickness. The thinnest sheets were too thin to show the common colors of interference in thin films, but the thicker sheets showed these colors. With the

* *Zeit. f. Phys.*, **60**, 741, 1930.

† *Naturwiss.*, **18**, 706, 1930.

‡ *Proc. Roy. Soc.*, **119**, 663, 1928.

§ *Jap. Jour. Phys.*, **5**, 83, 1928.

first apparatus used, the electron stream was less homogeneous than was desired. A magnetic field was then provided which deflected the electrons issuing from the first aperture along a portion of the arc of a circle, and a second aperture was provided along this path for the purpose of selecting a more nearly homogeneous beam. The range of bombarding potentials extended from 10,000 to 85,000 volts.

Figures 244 and 245 show typical patterns obtained by Kikuchi when using respectively one of the thinnest sheets of mica and one of the thicker

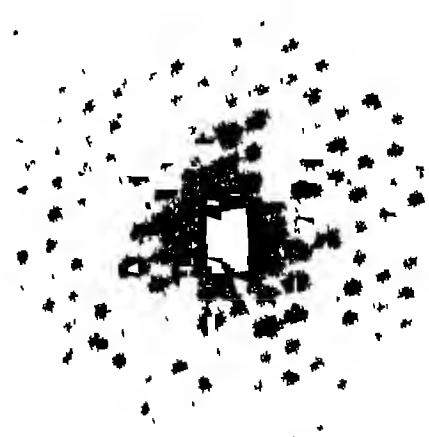


FIG. 244.—Kikuchi pattern from mica—very thin sheet.

FIG. 245.—Kikuchi pattern from mica—thicker sheet.

ones. Four features are to be noted in these patterns and these require interpretation:

In Figure 244, thinnest sheet:
  Netlike array of dark spots [N-pattern]
In Figure 245, thicker sheet:
  Concentric rings
  Laue spots on the rings [L-pattern]
  Straight dark lines, and, parallel to the more prominent of these, faint light lines. Each light line lies closer to the center of the pattern than does the corresponding dark line. They occur in pairs [P-pattern]

Mica is a monoclinic prismatic crystal. According to X-ray examination, the "unit cell" (Fig. 246, upper portion) has dimensions:

$$a=5.17\ \text{Å};\quad b=8.96\ \text{A};\quad c=20.5\ \text{Å};\quad \beta=84°10'.$$

The cleavage plane, the plane of the sheet of mica, is parallel to the base of the unit cell, that is, horizontal in the upper figure. The base of the unit cell is a rectangle (see lower figure). There is a lattice point at each corner of the rectangle and one at the center. A molecule of $KH_2Al_3(SiO_4)_3$ "centers"

about each lattice point. The detail of the arrangement of the atoms does not concern us. The relative magnitudes of a and b are such that the *semi-diagonal* of the rectangle, which equals $\sqrt{(a/2)^2+(b/2)^2}$, turns out to be numerically equal to a. Thus the lattice points in the net planes parallel to the plane of the sheet of mica form equilateral triangles. These net planes are themselves separated by 20.4 Å.

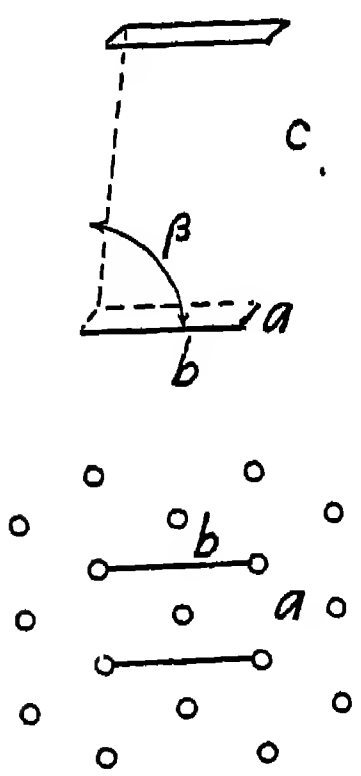


FIG. 246.—Above, unit cell for mica; below, lattice points in cleavage plane.

In the Laue diffraction of *X-rays* only spots occur, and these occur only when the radiation of the source includes certain discrete wave-lengths. For this reason a source of X-rays yielding a continuous spectrum, "white radiation," is used.

In the formation of the Kikuchi electron-diffraction patterns we must have also to do with some form of Laue diffraction because the electrons are diffracted *through* the mica. Yet these patterns appear at any bombarding potential—at any wave-length chosen at random, within certain wide limits. The question arises why any pattern at all is formed when the incident radiation is monochromatic with the wave-length chosen at random, and why the patterns reveal the three features above enumerated in addition to Laue spots.

The N-pattern (Fig. 244), obtained with the thinnest sheet of mica, is, at every bombarding potential, in nature and size that which is to be expected from a single net of lattice points arranged in equilateral triangles of the size of those which obtain in the plane of the sheet of mica. This is on the supposition that the De Broglie relation gives the wave-length associated with the electron beam correctly. Thus, this Kikuchi pattern, whether we can interpret it or not, yields further verification of the De Broglie relation, if indeed this were needed. The pattern is that of a single plane net, as distinguished from that of a space lattice, yet instead of a single net plane there are numerous identical net planes in the sheet of mica.

The pattern is practically the same whether the sheet of mica be perpendicular to the incident stream or be tilted several degrees out of perpendicularity.

Let us consider again, as we did in chapter 8, section 16, what happens when radiation is diffracted by a single row of lattice particles lying *along* the direction of the incident beam. The diffraction of zero order yields a point on the photographic plate which lies on the continuation of the incident beam through the crystal. The first order of diffraction yields a circle about the former point as center. The second order of diffraction yields a

larger circle, etc. This applies when there are many lattice points in the row lying along the incident beam. But when there are only, let us say, ten lattice points in the row, the maximum of zero order spreads out into a circular patch of which the size, when calculated, is surprisingly great. Simultaneously the circles of the first and second order, etc., become appreciably broad rings.

When a crystal is *sufficiently* thin, a pattern of the N-type may be formed entirely within the circular patch of zero order. In this case the black spots of the pattern represent simply the plane-grating beams which, in accord with calculation, fall within this maximum. This aspect of the explanation of the N-pattern was first definitely given by W. L. Bragg and F. Kirchner, and Kirchner has since shown that this explanation actually applies to the pattern formed by a film of celluloid or of collodion when this has a thickness of only 10^{-6} or 10^{-7} cm and contains single crystals of sufficient size to permit performing the experiment.*

Kikuchi's thinnest sheets of mica are thicker than 10^{-6} cm; they are probably of the order of 10^{-5} cm. Thus, with the net planes separated by 20.4 Å, there are about fifty identical net planes. However, the central patch of zero order is of considerably greater extent than would be expected for a row of lattice particles fifty in number. A presumption is thus established that each electron wave packet is diffracted by comparatively few particles of the row and that the remaining particles which are present do not contribute to reducing the angular size of the maximum. Effects due to multiple scattering also enter here, however.

There can be no doubt that a considerable amount of multiple scattering occurs. The low penetrating power of an electron beam as compared with an X-ray beam, even a soft X-ray beam, is evidence for the relatively great influence of each layer of atoms of the crystal upon incident electrons. And multiple scattering is a natural concomitant of this great influence and rapid extinction.

There is another factor besides the one first emphasized by Bragg and Kirchner which may play an important rôle in the formation of the N-pattern; W. Linnik† has shown that when monochromatic *X-rays* are diffracted by a sheet of mica which has been previously heated to redness, and thus caused to disintegrate into many thin laminae, then a pattern approaching the N-type is obtained. It cannot be properly maintained in this case that the mica has disintegrated to such a degree that each layer of lattice particles acts independently of other layers, or that groups of two or three layers

* See W. L. Bragg and F. Kirchner, *Nature*, **127**, 738, 1931; F. Kirchner, *Zeit. f. Phys.*, **76**, 576, 1932; and *Ergebnisse d. Ex. Naturwiss.*, **11**, 70, 1932.

† *Nature*, **123**, 604, 1929.

act independently of other similar groups. What is probably the correct explanation has been given by W. L. Bragg* and by S. B. Hendricks.† The laminae are tilted out of their original plane by various small angles. This tilting takes place for various laminae about various axes, but all of the axes of tilt lie in the plane of the sheet. There is no turning of the laminae over each other about an axis perpendicular to the plane of the sheet—there is no rotation in azimuth. The condition then approximates that of the powder spectrograph in so far as tilting out of the original plane is concerned, but remains that of the Laue experiment in so far as the laminae have not been turned within their own plane. Under this condition one would expect to find a pattern of the N-type, which is found, although the diffraction is due to numerous space lattices. Bragg supports this view by a subsidiary experiment of his own in which he first, by way of check, obtains a Laue pattern, using "white" X-rays and and an ordinary sheet of mica; not one previously heated, like Linnik. He then uses monochromatic X-rays with the same ordinary sheet of mica, but during the exposure "rocks" the mica to and fro over an angle of several degrees in such a manner that the normal to the sheet passes a number of times through all possible positions within a certain solid angle described about the original position. He *then obtains an N-pattern.*

The bearing which Linnik's and Bragg's experiments with X-rays have upon the explanation of Kikuchi's N-pattern obtained with electrons is this: Perhaps the sheet of mica is locally distorted due to the electron bombardment and should be considered as made of numerous constituent laminae which are tilted at slight angles with reference to each other but not turned in their own plane. Because of this tilting the laminae may produce a pattern of the type to be expected from a single net, although the diffraction really arises from numerous tilted space lattices.

Turning now to the features exhibited by Kikuchi's thicker sheets of mica: The thicker sheets are of the order of 10^{-4} cm in thickness and thus contain about 500 identical net planes. To obtain the pattern shown in Figure 245 the sheet of mica is tilted out of the plane perpendicular to the incident beam in such a manner that a row of lattice points lies along the beam. This requires a tilt of $(90° - \beta)$, or about 6°.

The concentric rings are found to lie, according to calculation based upon the De Broglie relation, at the positions of the diffraction maxima due to a row of lattice particles spaced at the distance which obtains in mica. The dark spots on these rings are the electron-wave analogues of Laue spots. The reason why there are always several such spots present at any bombarding potential is no doubt connected with the great effect of each layer of

* *Ibid.*, **124**, 125, 1929. † *Zeit. f. Kristall.*, **71**, 269, 1929.

atoms of the crystal. This great effect means that relatively few layers take part in the diffraction of any given wave packet. Consequently, the "resolving power" is low and then in turn the wave-length at which a given spot appears is not very sharply defined. Hence, at any given bombarding potential, there are always some spots which are either growing, or are at full development, or are fading. Multiple scattering, which certainly takes place, probably also plays a rôle in causing the spots to be present at any bombarding potential and in causing the continuous blackening along the concentric rings.

The P-pattern, consisting of pairs of black and white lines, is definitely attributable to multiple scattering. A good proportion of the electrons lose little or no energy in this scattering, so that the wave-length associated with these electrons will be practically unaltered. The scattering causes the beam to diverge within the crystal. Referring to Figure 247, let M represent the sheet of mica and i the incident beam. Within the crystal the beam diverges due to scattering. Let the broken line N be parallel to an important set of net planes, and let the arrow s represent the direction in the divergent beam which is such that reflection in these net planes will take place according to Bragg's law. The Bragg reflected beam takes the direction r and gives rise to a practically straight dark line upon the photographic plate, whereas in the direction e the beam suffers extinction, and hence a white line lies in this direction. The dark and white lines thus represent a sort of electron-wave analogue of the reflection and extinction lines of Rutherford and Andrade, discussed in chapter 8, section 8. In the present case there is also some compensating effect by scattering from the under surface of the set of planes N, which will give rise to a dark line at e and a white line at r, but the compensating effect must be slight because the beam scattered from the under side of the planes must have a much lower original intensity, since it makes an appreciably larger angle with the incident beam i. Kikuchi applies the explanation in detail, accounting for the precise positions in which the black and the white lines are found. He also accounts for the amount by which the lines shift in position as the crystal is tilted.

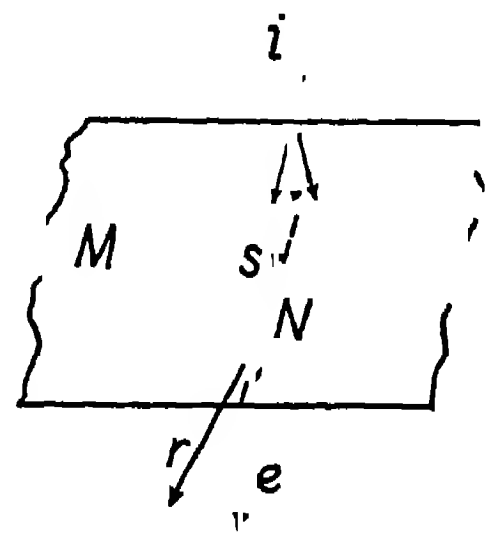


FIG. 247.—Manner of origin of Kikuchi lines.

14. The Diffraction of Fast Electrons by Thick Single Crystals.—The method of electron diffraction which we are now about to consider was tried soon after the appearance of the first researches of Davisson and Germer and of Thomson, but real success was attained only after the technique of electron diffraction had been developed far beyond that of the early days.

Referring to Figure 248, let i represent an incident beam of fast electrons and XY a crystal. Let us suppose for simplicity that i is incident directly over the row of lattice particles X. The beam may be incident either in the

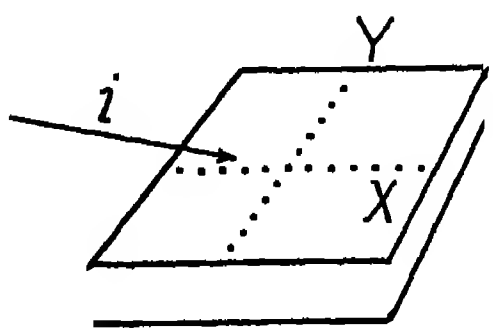


FIG. 248

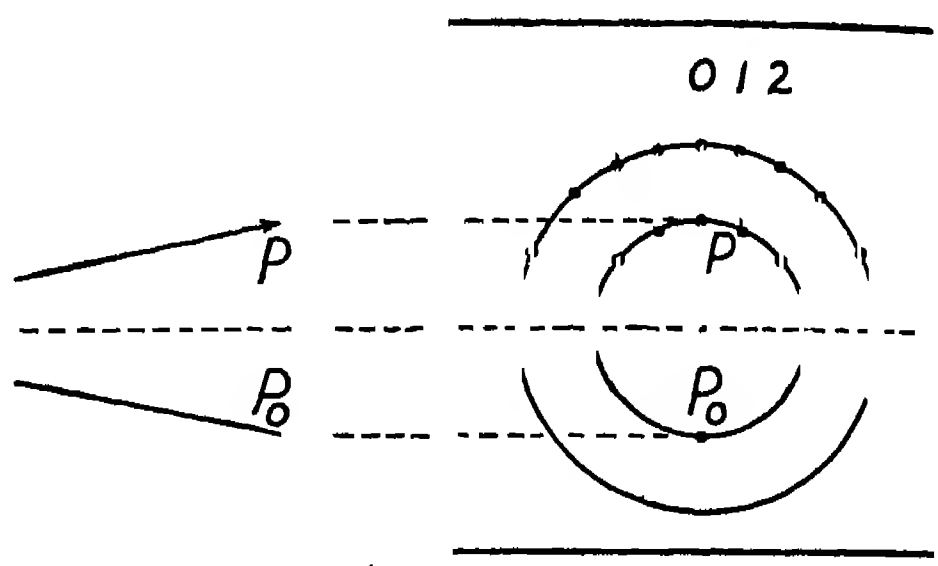


FIG. 249

grazing direction or at a glancing angle of several degrees. Let P_0 represent the continuation of the direction i beyond the crystal, and let P represent that direction lying over the row of lattice points X for which the angle of reflection is equal to the angle of incidence. For the row of lattice particles X the zero order of diffraction is a circular cone about X as axis which contains the lines P_0 and P and intersects the plane of the figure along these lines. This cone, of zero order, will intersect a photographic plate erected at the position P_0P, in the circle P_0P of Figure 249. The cone representing the first order of diffraction will now intersect the photographic plate in a larger circle such as the one indicated, and higher orders will intersect the plate in still larger circles.

The row of lattice particles Y yields cones about Y as axis which are of wide angle for low orders of diffraction and hence intersect the photographic plate in the practically straight vertical lines 1, 2, etc.

The surface layer of atoms acting alone as a single plane net would give rise to a pattern of dark spots at the various intersections of the circles with the vertical lines 1, 2, etc. This applies for a simple rectangular lattice. With a face-centered lattice, certain of the spots would not be expected to appear. In actual experiment the pattern expected for the given lattice employed is found, under suitable conditions, indicating that under these conditions the surface layer of atoms predominates in causing the diffraction. Under suitably modified conditions the Kikuchi pattern of black and white lines appears. Under again different conditions a pattern of spots arranged in rectangular array is obtained. It would take us too far afield to discuss these various cases in detail.

The method of electron diffraction which is under discussion has proved itself valuable in investigating the nature of the surfaces of crystals and the

nature of surface layers of foreign atoms deposited on originally clean crystal surfaces. The method is also valuable in investigating the refraction of electron waves and in deducing values of inner potentials. This method gives clearly interpretable results, whereas in the bombardment of crystals by slow electrons many anomalous weak beams appear which yet require interpretation, and other anomalies likewise present themselves. The method of bombarding single crystals with fast electrons promises to become increasingly important. The reader is referred to the sources listed in the footnote for further information.*

15. The Scattering of Electrons by Gases and Liquids.—The investigations involving gases have followed three principal lines:

1. *The application of Thomson's method.*—In place of the thin solid film which is ordinarily used in Thomson's method, a jet of gas issuing from the end of a fine nozzle passes perpendicularly across the electron beam. After passing across this, the stream of gas is at once condensed upon a surface cooled by liquid air, or is removed by rapid pumping, in order that the vacuum which obtains in the apparatus shall not be impaired. A photographic plate is placed perpendicular to the undeviated beam.

This method is most suitable for the investigation of the scattering or diffraction due to polyatomic gases containing two or more atoms of at least fairly high atomic number per molecule. With such gases, interference of the disturbances scattered by the individual atoms gives rise to a pattern consisting of several broad diffraction halos. However, as photometric tracings taken from the photographic plates show, the blackening does not attain a maximum at each halo, but merely undergoes a marked variation of gradient.

Among the numerous gases which have been investigated by this method are, for example, those of the series CCl_4, $CHCl_3$, CH_2Cl_2, etc.

The amount of scattering to be attributed to each atom of the molecule is proportional to the square of the atomic number. Hence the heavy atoms do most of the scattering. Or, in other words, taking the case of $C_2H_2Cl_2$, the formation of the pattern is almost entirely due to the interference of the electron waves scattered by the two chlorine atoms. Bragg's law applies, as in the scattering of X-rays by gases (see chap. 8, sec. 25 in this connection), and it is found with electron bombardment, as with X-rays, that the "cis" form of $C_2H_2Cl_2$, in which the chlorine atoms are closer together, yields a

* G. P. Thomson, *Proc. Roy. Soc.*, **133**, 1, 1931; Kirchner and Raether, *Phys. Zeit.*, **33**, 510, 1932; H. Raether, *Zeit. f. Phys.*, **78**, 527, 1932; K. Shinohara, *Tokyo Sci. Papers*, **18**, 223 and 315, and **20**, 39; T. Yamaguti, *Proc. Phys.-Math. Soc. Jap.*, **12**, 203, 1930, and **14**, 1 and 57, 1932; summarizing account by F. Kirchner in *Ergebnisse d. Ex. Naturwiss.*, **11**, 80, 1932.

pattern of more widely spaced halos than the "trans" form, in which the chlorine atoms are farther apart.

The method of electron diffraction which is under discussion requires an exposure time of less than a second, and herein it presents a great advantage over investigations on the diffraction of X-rays by gases, which require exposures of about twenty hours.

The method was first applied by Mark and Wierl in 1930, and has since been applied especially by Wierl but also by others.*

Bombarding speeds of some fifty to a hundred kilovolts are used. The angles of scattering involved in the formation of the diffraction halos are small ones.

2. *Scattering at large angles.*—The second type of investigation which we shall discuss involves determination of the amount of scattering as a function of the angle of scattering from zero to as nearly 180° as it is mechanically possible to go. The scattering due to monatomic gases has received the most attention in investigations of this type because, for such gases, it is possible to deduce theoretically a function for the expected variation of scattering with angle. Opportunity for making comparisons between theory and experiment is thus presented. Helium has been the most thoroughly investigated, but argon, neon, and Hg vapor have likewise been studied. Some diatomic gases—for example, hydrogen and nitrogen—have also been investigated experimentally, and likewise a few polyatomic gases—for example, carbon dioxide and methane.

The experimental method consists in causing an electron gun of the usual type to project a homogeneous beam of electrons into a body of the gas under investigation, the gas being at such low pressure that multiple scattering is negligibly improbable. The number of electrons scattered to any given angle is then determined by allowing the electrons to enter a Faraday-box collector which is directed toward a point about a centimeter in front of the muzzle of the gun, and is capable of rotation about this point as center. Slow electrons are used for bombardment. Curves are usually taken at each of several bombarding potentials. The bombarding potentials which have been used range from 4 volts to 800 volts. Frequently, measurements have been confined to those electrons which have been scattered elastically—that is, with little or no loss of velocity—but in some instances measurements have also been made upon those electrons which have lost given amounts of energy.

Beginning at zero angle, the number of electrons scattered per unit solid angle always falls off rapidly at first with increasing angle of scattering, and

* *Naturwiss.*, **18**, 205, 1930; R. Wierl, *Phys. Zeit.*, **31**, 366, 1930; *Ann. d. Phys.*, **8**, 521, 1931, and **13**, 453, 1932; also, e.g., Brockway and Pauling, *Proc. Nat. Acad.*, **19**, 68, 1933.

then falls off more slowly as the angle is further increased to, let us say, 20° or 30°. Thereafter—depending, however, upon the gas which is under investigation and the bombarding potential—there may be one or two low maxima at large angles; and, moreover, when the curve is followed to as near 180° as possible, it often rises at the end, indicating another low maximum of scattering in the directly backward direction.

While the number of electrons scattered per unit solid angle on the whole decreases markedly in passing from a small angle of scattering to a large one, the total number of electrons scattered between, let us say, the angles of 10° and 20° is much smaller than the total number scattered between 80° and 90°. This is because a zone described between the limits 10° and 20° upon the reference sphere, has a much smaller area, or, in other words, subtends a much smaller solid angle than a zone described between 80° and 90°. Accordingly, while a curve showing the scattering per unit solid angle has a high maximum at zero angle, a curve showing the total scattering between each of equal angular intervals, begins at zero, at zero angle, and subsequently attains one or more maxima.

The first investigation of the type under discussion was made upon helium. A first report by E. G. Dymond, in the year 1927, later proved to have given entirely erroneous results. A second report, in which correct results were given, was made by Dymond and Watson in 1929. Since then numerous investigations of the same general type have been made.*

3. *Investigation of total scattering.*—When an originally homogeneous and well-defined beam of electrons passes through a gas, one may determine the total number of electrons which are scattered out of the beam by a column of the gas of given length. No information is then obtained in regard to the angles at which the electrons are scattered or whether they are scattered elastically or inelastically, and with what loss of energy if any. But the results are nevertheless of great interest. When the total number of electrons scattered out of the beam is plotted as ordinate, against the velocity of the incident stream as abscissa, the curve obtained shows most unexpected features. These features lead to corresponding surprising conclusions regarding the manner in which the effective area of cross-section of the atoms or molecules of the gas under bombardment varies with the speed of the bombarding electrons.

* See, e.g., the following, where further references will also be found: F. L. Arnot, *Proc. Roy. Soc.*, **130**, 655, 1931; Tate and Palmer, *Phys. Rev.*, **40**, 731, 1932; Hughes, McMillen, and Webb, *ibid.*, **41**, 154, 1932; Ramsauer and Kollath, *Annal. d. Phys.*, **12**, 837, 1932; Jordan and Brode, *Phys. Rev.*, **43**, 112, 1933; Hughes and McMillen, *ibid.*, **44**, 20, 1933.

Theoretical treatments have been given principally by N. F. Mott, *Proc. Roy. Soc.*, **127**, 658, 1930, and Massey and Mohr, *ibid.*, **136**, 289, 1932.

Electrons having velocities of above a thousand volts pass through the gas atoms readily. For such electrons the effective area of cross-section, A_e, of the atoms or molecules is a small fraction of the area of cross-section as calculated from the kinetic theory of gases, which we shall call A_k. As the speed of the incident electron stream decreases, A_e gradually increases however, and at 50 volts or thereabouts, depending upon the gas in question, attains a value equal to A_k. As the electron speed is further decreased, the results become striking. Those for Krypton are typical: The effective area, A_e, continues to rise until at 11 volts it attains a maximum of four times A_k. Hereafter it falls rapidly until at .6 volt it passes through a minimum of one-seventeenth of A_k. Thereafter A_e again rises, and at the lowest potential at which measurements are possible, some .2 volt, it is rising rapidly.

The foregoing striking variation of area of cross-section with bombarding potential began to be revealed in about the year 1921 by the work of J. S. Townsend and his followers, especially V. A. Bailey. Their work did not however, attract the general attention which it merited. The effect was independently and more forcefully demonstrated by a method differing from that which Townsend used, by C. Ramsauer and his followers, especially R. Kollath, and is called the "Ramsauer effect" or "Townsend-Ramsauer effect."*

The diffraction of electrons by *liquids*—applying Thomson's method for thin solid films—has been carried out by L. R. Maxwell.† The application is limited to liquids which have a very low vapor pressure and in addition have suitable surface tension and viscosity to permit forming a thin film in a wire loop. The film must be thin enough to transmit the electrons. Several heavy oils were investigated, and patterns consisting of a central spot and two or three surrounding halos have been photographed.

Buehl and Rupp‡ have used a method of diffracting electrons by liquids which presents a certain analogy with the experiments of Davisson and Germer on the Bragg reflection of electrons described in section 10 of the present chapter.

16. The Polarization of Electron Waves.—Soon after the wave nature of electrons had been conclusively demonstrated by the experiments of Davisson and Germer and those of Thomson, various investigators sought to determine whether electron waves could be polarized. The method of attack closely paralleled the method by which the polarization of light is detected

* For further information regarding the effect itself and the development of its study see, e.g., Minkowski and Sponer, *Ergebnisse d. Ex. Naturwiss.*, **3**, 67, 1924; Broese and Saayman, *Annal. d. Phys.*, **5**, 797, 1930; R. Kollath, *Phys. Zeit.*, **31**, 985, 1930; J. S. Townsend, *Proc. Roy. Soc.*, **134**, 352, 1931.

† *Phys. Rev.*, **44**, 73, 1933. ‡ *Zeit. f. Phys.*, **67**, 572, 1931.

by two reflections. The most carefully executed of the early experiments were performed by Davisson and Germer.* In their work an electron gun furnished a beam of electrons, i (Fig. 250), which was incident upon a target, T_1, which consisted of a single crystal of nickel, and the once-reflected beam was again reflected by a similar target, T_2. The twice-reflected beam, A, then entered a receptor, R. The velocity of the original beam was adjusted to one of the discrete values which yield a Bragg reflected beam; these discrete values, it will be recalled, depend upon the value of the angle of incidence, and this angle was 45° in the experiments in question. Only electrons which had retained their full speed after the two reflections were admitted into the collector. Under these various conditions of experiment the prospect of detecting polarization was considered to be the most favorable. The receptor and second target were rotated as a unit about the line T_1T_2 as axis, the rotation being performed step by step, and the intensity was measured at each step. If the polarization phenomena of electron waves closely paralleled those of light-waves, a maximum of intensity should occur in the direction A, which we shall denote by the azimuth $\psi=0$; a minimum of intensity should occur in the direction perpendicular to the figure, azimuth $\psi=90°$; a second maximum should occur in the direction B, azimuth $\psi=180°$; and a second minimum should occur in the azimuth $\psi=270°$. Such a variation was not found, however. The intensity was, in fact, found to be the same in all azimuths to within the error of measurement, which was one-half of 1 per cent. Various discrete velocities of the incident electrons were used, up to a value of 120 volts.

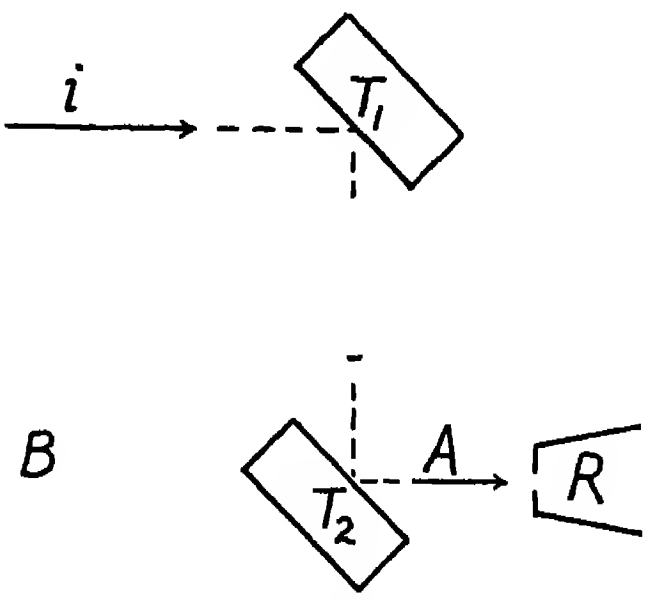


FIG. 250

While the above-mentioned experiments were in progress, N. F. Mott† developed a theoretical treatment of the polarization of electron waves by scattering, based upon the Dirac relativistic wave equation for the motion of the electron. Mott considers the case in which an electron is deflected in succession by two isolated stripped atomic nuclei—that is, the two targets T_1 and T_2 of our figure are supposedly each replaced by an isolated nucleus. The theory of the problem involves the "spin" or "magnetic moment" of the electron regarded as a particle. The original beam is supposedly unpolarized, that is, the spin axes are supposedly oriented at random. After the first scattering the spin axes will, however, show a preferred orientation,

* *Phys. Rev.*, **33**, 760, 1929.

† *Proc. Roy. Soc.*, **124**, 425, 1929; see also a later paper, *ibid.*, **135**, 429, 1932.

to a degree depending upon conditions which will be mentioned later. The once-scattered beam will thus be partially polarized, and it should be possible to detect this polarization by allowing a second scattering to take place and measuring the number of electrons scattered in various azimuths.

Mott's theory indicates, however, that the polarization effect to be expected for electron waves will differ in a fundamental respect from the effect which exists for light-waves. For an electron the spin axis has a definite direction and also a definite "sense" along this direction, whereas plane-polarized light has a definite direction or azimuth of polarization but not a sense of direction along this azimuth. One sense of direction along this azimuth is not distinguishable from the opposite sense. Because of this difference, an electron beam polarized by double scattering will show only one maximum and one minimum per revolution in azimuth, and not two maxima and two minima per revolution, as is the case with plane-polarized light. An electron stream which has been twice scattered by isolated atomic nuclei should have a maximum of intensity in the azimuth which we have designated by 180° and a minimum in the azimuth 0°, and the azimuths 90° and 270° should have intermediate values.

Theory further indicates that the polarization effect should increase with the square of the atomic number of the scattering nuclei and should increase with rise in velocity of the incident stream of electrons until a velocity of about .7 of the velocity of light has been reached. Then, passing through a maximum, the effect should again become smaller.

The amount of polarization which is to be expected may be numerically evaluated, and such evaluation shows that when velocities of the incident stream are used as low as those used by Davisson and Germer, up to only 120 volts, the amount of polarization would be indetectable even with the heaviest possible scattering nuclei. There is consequently now good reason to comprehend why the attempt of Davisson and Germer to detect polarization effects led to negative results.

Obviously one should work with scattering atoms of high atomic number and use high voltages for accelerating the electrons. But even when these conditions are fulfilled, the existence of a polarizing effect is difficult to establish, and it is still more difficult to make any quantitative comparison between theory and experiment. It is necessary to use accelerating potentials of 50,000 volts or more, and when such voltages are used, stray effects are likely to mask the effect sought for. Furthermore, the most reliable theoretical treatment—namely, that of Mott—calls for isolated stripped nuclei, whereas actually the scattering atoms are surrounded by electrons and the atoms form crystalline aggregates instead of being isolated. Theoretical treatments based upon *crystals* used as targets have also been given,

but they lead to less definite and reliable conclusions than the case treated by Mott.

As for the experimental results when high-voltage electrons or high-speed β-particles are used, several investigators believe that they have now found an effect of polarization. Among these are to be mentioned especially C. T. Chase, and Rupp and Szilard, and E. G. Dymond. But conclusions must be drawn with some reserve, as the experimenters themselves for the most part realize.

It is still an open question whether or not electron beams show polarization effects under given realizable experimental conditions. We shall consequently not enter into further details; suffice it to refer to two articles where summarized discussion and complete references to the existing literature may be found.*

EXPERIMENTS WITH ATOMS, MOLECULES, AND POSITIVE IONS (Secs. 17–21)

Following the discovery of the wave nature of electrons, several investigators sought experimental evidence for the wave nature of atoms, or molecules, or positive ions, by bombarding crystals with one kind or another of these particles.

17. General Considerations and Technique.—With positive ions the electrical methods which were in use for producing and receiving streams of electrons were available to be so modified as might be necessary. Moreover, positive ions having speeds which obtain in a discharge tube produce an effect upon a photographic plate. From these circumstances one might infer that success would have come first with positive ions rather than with neutral atoms or molecules. With the latter, electrical methods are not applicable, and, also, streams of these particles do not affect the photographic plate. Hence entirely new technique needed to be developed. Nevertheless, real success was first had by bombarding crystals with neutral particles.

Marvelous technique was developed by O. Stern and his collaborators, F. Knauer, I. Estermann, and R. Frisch, for producing and receiving streams of helium atoms and hydrogen molecules and then using streams of these light gases in experiments on the diffraction of material particles. This technique we shall now outline:†

The atom or molecule "gun," G (Fig. 251), consists simply of a small chamber provided with a slit S_1. Gas particles issue from the chamber

* G. O. Langstroth, *Proc. Roy. Soc.*, **136**, 558, 1932; E. G. Dymond, *ibid.*, p. 638.

† For further details than can be here given see Knauer and Stern, *Zeit. f. Phys.*, **53**, 766 and 779, 1929; Estermann and Stern, *ibid.*, **61**, 95, 1930; Estermann, Frisch, and Stern, *ibid.*, **73**, 348, 1932.

through the slit S_1 into a surrounding inclosure which is evacuated by continuous rapid pumping. A fresh supply of gas is continually fed from a large reservoir through the tube t. The particles leaving the gun have the velocities of the Maxwell temperature distribution appropriate to the temperature of the gun, which may, by suitable arrangements, be given a wide range of values. The stream of particles is heterogeneous in velocity and is hence equivalent to "white radiation." The stream which issues from the slit S_1 is, moreover, divergent. A second slit S_2 selects a nearly parallel beam out of the original one, and this nearly parallel stream is then incident upon the under surface of the crystal C. Speaking from a general standpoint, we must now suppose that beams emerge from the crystal in various directions. The crystal may be rotated about the axis A for the purpose of presenting various azimuths of the face under bombardment to the incident beam. A slight universal tilting of the crystal is provided, moreover, for the purpose of adjustment and for another purpose which will appear later.

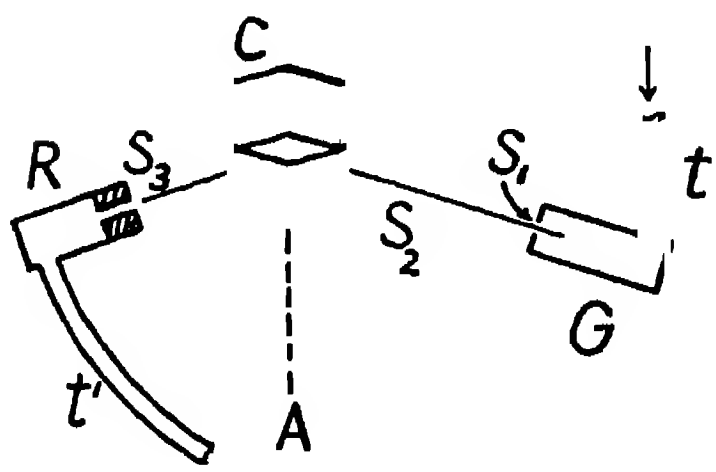


FIG. 251.—Apparatus of Estermann and Stern.

The slits S_1 and S_2, and also S_3, as yet to be spoken of, all lie in the plane of the diagram. That is, the apertures shown in the diagram represent in each case the *length* of the slit, and not the *width*.

Those particles which leave the point of bombardment of the crystal in the appropriate direction pass through the slit S_3 into the receptor R. The principle upon which the receptor operates will be explained shortly.

A second inclosure surrounds the crystal and receptor, and this inclosure is also kept highly evacuated by continuous and rapid pumping. This inclosure can be, and must moreover be, maintained at a higher vacuum than can well be attained in the inclosure surrounding the gun. The vacuum which can be attained in the inclosure surrounding the gun is evidently limited by the fact that the entire divergent stream issuing from the slit S_1 feeds into this inclosure, whereas the vacuum which can be attained in the inclosure surrounding the crystal and receptor is limited only by the much lesser stream which passes through the slit S_2.

We shall now explain the principle upon which the receptor operates: The chamber of the receptor R connects through a tube t' to a small glass vessel which is entirely closed except where the communicating tube t' enters. This closed vessel contains the essential element of an extremely sensitive pressure-measuring device which we shall describe presently. The entire receptor system—that is, chamber R, communicating tube t', and the

vessel containing the sensitive pressure-measuring device—thus constitutes a system which is closed everywhere except for the slit S_3. The stream of particles coming from the point of bombardment of the crystal enters the slit S_3 and builds up a pressure in the receptor system until particles escape out through the slit S_3 at a rate equal to that at which particles are entering this slit. The maximum pressure built up in the receptor system is thus a measure of the rate at which particles are entering through the slit S_3 or is, in other words, a measure of the intensity of the beam coming from the bombarded area of the crystal.

The essential element of the delicate pressure-measuring device or "manometer" is a fine sealed-in wire which is heated by a constant electric current. The temperature attained by this wire, and hence its electrical resistance, is a function of the number of particles in the vessel which conduct heat away from the wire, or is, in other words, a function of the pressure. The intensity of the beam coming from the crystal is thus measured by measuring the pressure which is built up in the receptor system, which is measured by determining the amount by which the resistance of the fine heated wire changes. The amount of this change is registered by the deflection of a galvanometer in a Wheatstone bridge circuit.

A manometer of the type above in question is known after its inventor as a "Pirani hot-wire gauge." Especial designs of this form of gauge were developed by Knauer and Stern,* in order to attain a sensitivity sufficient to measure the minute changes of pressure which occur in experiments of the type which we are discussing.

The entire crystal spectrometer was widely modified in various stages of development and for various purposes. Generally speaking, however, the following details apply:

The length of arm of the spectrometer—that is, the distance from the center of the crystal to the slit S_3—is 15 mm.

The highest pressure which obtains in the entire apparatus is that in the gun, and this is considerably below 1 mm of mercury. The pressure in the gun must be maintained at a low value, otherwise the mean free path of the gas particles would be small and frequent collisions would occur just outside of the slit S_1. These collisions would result in the formation of a "gas cloud" hovering about the slit, and this cloud would act as a virtual source much broader than the slit itself. In this case the definition of the beam beyond the slit S_2 would be impaired. The mean free path of the particles, while within the gun, must be at least of the order of magnitude of the width of the slit S_1. Rather widely differing slit dimensions have been used from occasion to occasion, but in those experiments which are to be particu-

* *Loc. cit.*

larly discussed later, the slits S_1 and S_2 are to be thought of as being .2 mm wide and .5 mm long, and the slit S_3 somewhat wider and 1.5 mm long.

An important feature of the slit S_3 is that it has considerable "depth." Because of this depth it offers marked additional impedance to the escape of particles out through the slit, but offers no additional impedance to the entering stream in which the particles move straight down through the slit. The particles which start to escape have, on the other hand, random directions, and most of them strike the wall of the slit one or more times and may be turned back into the receptor. Accordingly, by having a slit of several millimeters depth, the pressure in the receptor is caused to build up to some ten or more times the value which would be attained if the slit were in a *thin* wall.

In the inclosure surrounding the gun the gas pressure is of the order of 10^{-4} mm, and in the inclosure surrounding the crystal and receptor it is about 1×10^{-5}. When the crystal is withdrawn and the apparatus is arranged, by suitable modification, so that the direct beam enters the receptor, the pressure within the receptor attains values of over 2×10^{-5} mm. The sensitive Pirani gauge is capable of detecting a pressure variation of 10^{-8} mm. Thus an intensity of one-one-thousandth of the direct beam may be measured.

The first achievement of various investigators in bombarding crystals with atoms or molecules consisted in showing that under proper conditions the streams were specularly reflected. The occurrence of specular beams constituted strong evidence for the wave nature of the bombarding particles, because incident atoms or molecules having merely the nature of particles would not be reflected in any well-defined direction any more than a stream of balls projected at a regular pile of balls would largely rebound in one direction; the direction in which each ball would rebound would depend upon the particular point at which it had made contact with a ball of the pile. Accordingly, the specular beams constituted evidence for the wave nature of the bombarding particles. But evidently the mere occurrence of specular beams does not give any information regarding the nature of the reflection or diffraction which is taking place, or lead to any method of determining the wave-length associated with the particles.

In the year 1928 T. H. Johnson succeeded in reflecting a beam of atomic hydrogen specularly from a cleavage face of natural rocksalt, and Ellett and Olson did the same for beams of cadmium and of mercury atoms. The work of these investigators will be more particularly described in sections 19 and 20. They used a technique differing from that above outlined.

In the year 1929 Knauer and Stern,* reflected beams of helium and molec-

* *Ibid.*

ular hydrogen specularly from natural rocksalt at various glancing angles, up to 50°, and also obtained meager indications of the existence of diffracted beams; but their technique was at that time insufficiently developed to reveal these beams definitely. They also searched, in vain, for the diffracted beams from a ruled grating.

With all particles other than electrons, the relatively large mass means that, at a given velocity, the associated wave-length will be correspondingly smaller. It will be recalled that according to the De Broglie relation any particle should have a wave-length given by the equation $\lambda = h/mv$. Even for the lightest of other particles, namely, the hydrogen atom, the wave-length will be only the 1/1850 part of that for an electron having the same speed, and all ensuing angles of diffraction will then be too small to resolve. However, the difficulty which thus arises because of larger mass is to a great extent offset by the fact that atomic and molecular streams produced in the general manner above outlined have much lower velocities than even the slowest electron streams which it is feasible to use. From the Maxwell distribution of velocities at room temperature, about 290° K, the associated De Broglie wave-length of maximum intensity, λ_M, is calculated to be .57 Å for helium and .80 Å for molecular hydrogen, or, in other words, the same respectively as for 460- and 235-volt electrons (cf. eq. [10′]).

18. The Diffraction of Helium Atoms and Hydrogen Molecules.—In the year 1930 Estermann and Stern* for the first time conclusively demonstrated the occurrence of diffracted beams of particles other than electrons. They bombarded crystals of rocksalt and of lithium fluoride, LiF, with beams of helium and molecular hydrogen. Synthetic crystals of lithium fluoride gave by far the best results and were therefore principally used. Lithium fluoride has the further advantage of having a smaller lattice constant than rock salt. It crystallizes in the same manner as rocksalt. The diffracted beams are, with either crystal, of the plane or crossed grating type. No space-lattice effect is evident; the layers of atoms, or rather ions, below the surface layer of the crystal, apparently play no rôle in the diffraction. Moreover, for some reason not thoroughly understood, the surface lattice behaves as though only ions of one kind—let us say the positive ones—were effective as diffracting centers. Figure 252 is a view looking down diagonally upon a cubic crystal. The circles represent, let us say, Li-ions of the bombarded face, and the crosses represent F-ions. Rows of ions parallel to the diagonals of the crystal face, or parallel to the X- and Y-axes indicated, contain only ions of one kind, only Li or only F. Suppose that two sets of wide grooves were cut into a plane surface, one set parallel to the X-axis and the other parallel to the Y-axis, the grooves being cut along the lines formed by the

* *Loc. cit.*

crosses. There would then result a rectangularly cross-ruled grating having elevations centering about the points occupied by the circles. The grating constant of this grating would be $d' = d/\sqrt{2}$. Experiment shows that the crystals behave like such a cross-ruled grating. The directions of ruling of the equivalent grating, or the principal directions of the lattice formed by *like* ions, are parallel to the *diagonals* of the crystal face. For LiF the length of edge of the unit cell is $d = 4.023$Å, and consequently $d' = d/\sqrt{2} = 2.845$ Å. We have then the problem of determining the nature of the diffraction pattern from a plane net of lattice points arrayed in squares of this dimension,

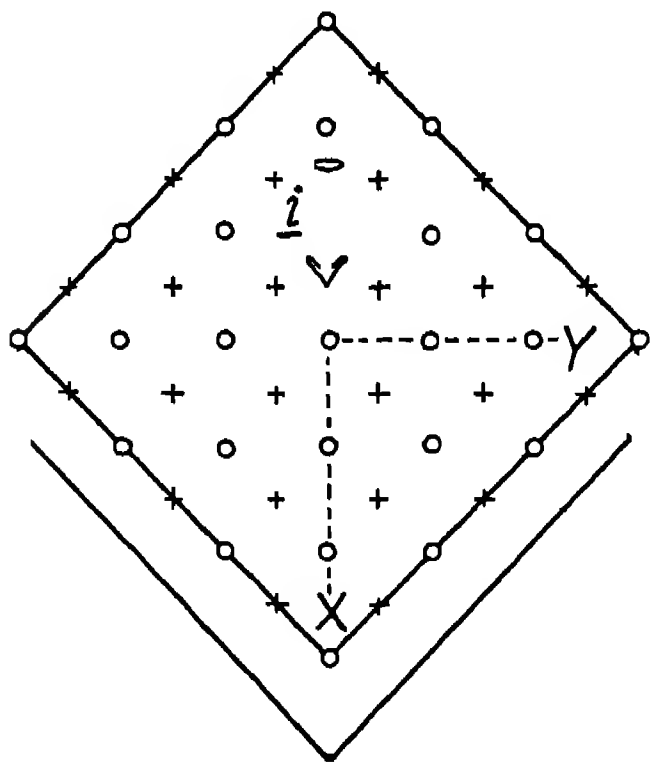


FIG. 252

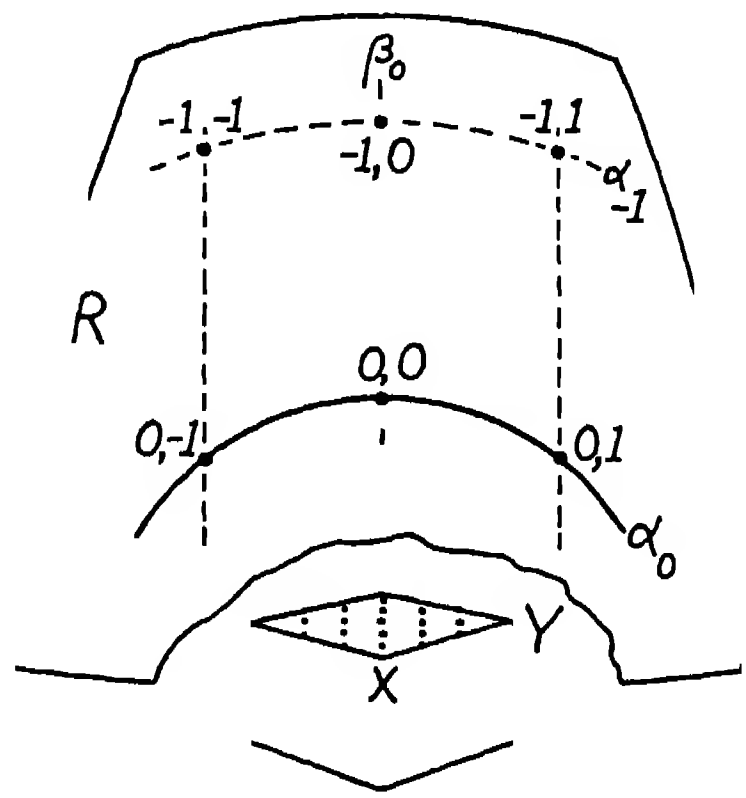


FIG. 253

the edges of the squares making angles of 45° with the edges of the crystal face.

For the purpose of discussion the incident beam will be considered as homogeneous, or monochromatic, instead of heterogeneous, or "white," as it actually is. Moreover, the cross-section of the beam will be considered as minutely circular instead of rectangular. Each diffracted image is thus to be thought of as a "spectrum dot."

Let us consider first the case when the incident beam represented by the arrow i (Fig. 252) lies over one of the diagonals of the crystal face. Call the angle made by the incident beam with the X-axis α_0. When the beam is incident in the manner supposed, α_0 is also the glancing angle made by the beam upon the crystal surface. Consider the row of lattice particles lying along the X-axis. The direction of diffraction from this row of particles generates a circular cone of angle α about the X-axis. There will be a cone for each spectral order, determined by the equation:

$$m_x\lambda = d'(\cos\alpha - \cos\alpha_0), \tag{17}$$

where m_x is the spectral order with regard to the X-axis. In Figure 253 R represents the contour of an arbitrary portion of a reference sphere taken about the point of bombardment of the crystal as center. The cones intersect the reference sphere in circles. The circle α_0 represents the intersection of the cone of order $m_x = 0$, for which $\alpha = \alpha_0$, and the circle α_{-1} represents the intersection of the cone of order $m_x = -1$ with this sphere. For the row of lattice particles lying along the Y-axis, the direction of diffraction generates circular cones, one for each order, about the Y-axis. These cones are determined by the equation:

$$m_y \lambda = d'(\cos \beta - \cos \beta_0), \tag{18}$$

where m_y is the spectral order with reference to the Y-axis, and β_0 and β are respectively the angles made by the directions of incidence and diffraction with this axis. These cones also intersect the reference sphere in circles, but the circles, when viewed as in the figure, appear as straight vertical lines. The central line, β_0, is the line of zero spectral order. The broken lines to the right and left represent the circles of order $+1$ and -1. The diffracted "spectral dots" appear at the intersections of the α-circles and the β-circles. At the intersection of the α_0-circle and the β_0-circle lies the spectrum dot of 0, 0 order, that is, the central image, or specularly reflected beam. To the right and left and somewhat lower lie the diffracted images of order 0, 1 and 0, -1. These are the images which are explored when the beam is incident, as supposed, over the diagonal of the crystal face. Other images—namely, -1, 0, -1, 1, and -1, -1—are also indicated in the drawing, but none of these is accessible to exploration when the apparatus is used according to the general plan which has thus far been outlined.

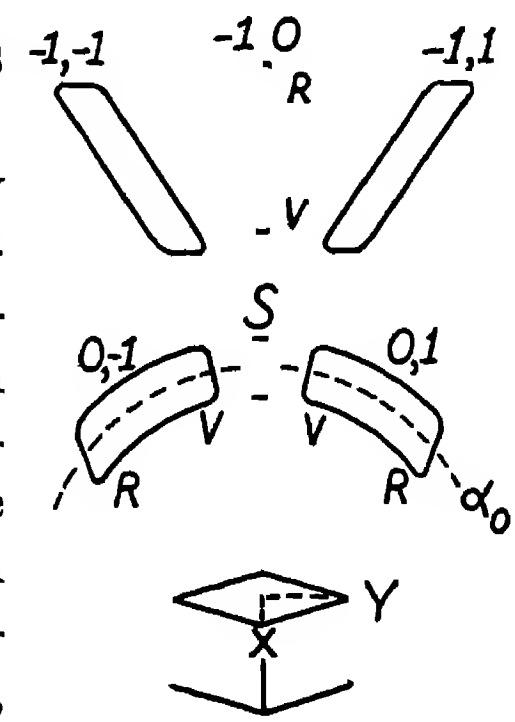


FIG. 254

Owing to the fact that the beam is in reality heterogeneous, or, as we say, "white," each diffracted image spreads out into a continuous spectrum extending toward and away from the central image. The resultant pattern is represented diagrammatically in Figure 254. In this drawing the appreciable height which the incident beam in reality has is taken into consideration. The central rectangle, S, represents the specular beam, the central image. To the right and left are the spectra of order 0, 1 and 0, -1, which are the ones explored. In accordance with optical terminology, the short wave-length end of each spectrum is lettered V, violet, and the long wave-length end, R, red; the wave-length of maximum intensity lies, roughly speaking,

midway between the ends. As was previously stated, when measurements are made, the receptor is rotated step by step about a vertical axis. When it is so rotated and all else remains fixed, it passes through the specular beam but misses the diffracted spectra for the most part. In order to make an exploration of both the specular and the diffracted beams, the crystal is slightly tilted about the Y-axis. At each setting of the receptor the angle of tilt of the crystal is adjusted to the optimum value, to yield maximum intensity. To be sure, the angle of incidence is slightly altered in doing this, but this does not matter; the diffraction pattern simply moves higher on the reference sphere. By proper tilting of the crystal the diffracted spectrum is always made to cross the slit of the receptor. In this way the pattern lying on the α_0-circle is explored in its entirety. Figure 255 shows the result of such an exploration. The beam is helium at room temperature, 290° K, incident at a glancing angle of $18\frac{1}{2}°$. The high central peak is the specular beam; the lateral peaks are the diffracted spectra. The arrows under the curve represent the positions at which the maxima are theoretically expected. The De Broglie wave-length of maximum intensity for a stream of helium at room temperature is, as has been previously stated, .57 Å. Applying equation (18) by inserting for d' its value 2.845 Å, and remembering that $\beta_0 = 90°$ and therefore $\cos \beta_0 = 0$, we have $\lambda_M/d' = .57/2.845 = .200 = \cos \beta$, whence $90 - \beta = 11\frac{1}{2}°$. This is the position at which the arrows are placed. The agreement between theory and experiment is all that can be desired. Similar curves have been obtained at other temperatures as well, at 100° K, at 180° K, and 580° K, for both helium and molecular hydrogen. Some minor discrepancies occur, but these can be plausibly accounted for.

FIG. 255.—Diffraction of a beam of helium, at 290° K, incident diagonally on the face of a crystal of LiF at a glancing angle of $18\frac{1}{2}°$ (Estermann and Stern).

The same apparatus was used by Estermann and Stern* for exploring another, different pair of diffracted images. The gun and receptor were

* *Ibid.*

made to point horizontally and the incident beam struck a *side* face of the crystal. Figure 256 represents the situation in perspective. The incident beam i makes a glancing angle θ_1 with the crystal surface. The plane of incidence is parallel to one pair of edges of the bombarded face. The diffracted images which lie in the plane of incidence are explored. They may be considered as arising from a vertically ruled grating of which the lines are formed by the vertical rows of *like* ions of the surface lattice. The grating constant of the equivalent grating is equal to the separation of the rows of like ions, namely, to $d/2 = 2.01$ Å. The specular beam is represented by S. To the right and left lie the diffracted spectra of order -1 and $+1$, respectively. Curves analogous to the one of Figure 255 were obtained. The theoretical positions of the maxima are readily computed by applying the plane-grating relation: $m\lambda = (d/2)\ (\cos\theta_2 - \cos\theta_1)$, where θ_2 is the glancing angle of the diffracted beam. The diffracted image on the right lies closer to the specular beam than does the one on the left. In fact, the image on the left can exist only when the glancing angle of the incident beam is sufficiently great.

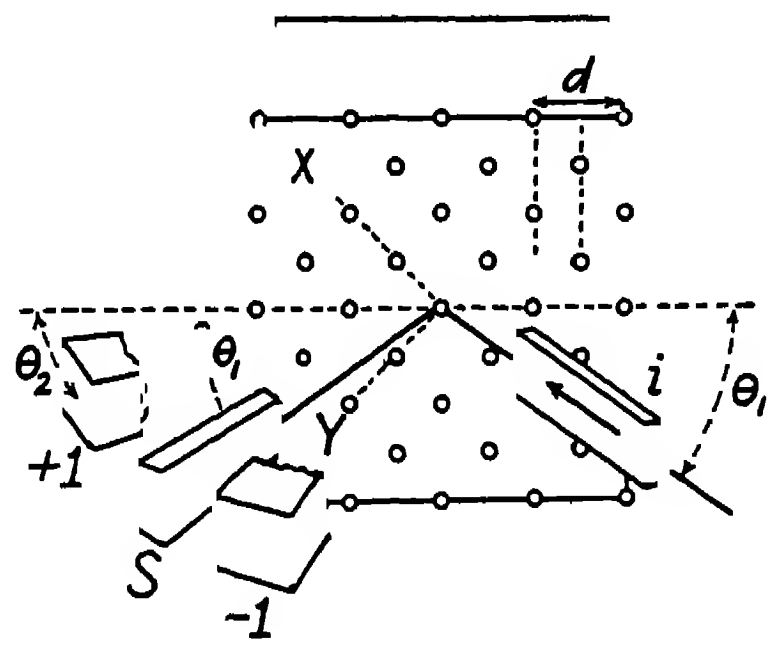


FIG. 256

Each diffracted image may be considered as occurring at the intersection of an α-cone and a β-cone generated about the X- and Y-axes. The axes are to be taken as shown, along the diagonals of the crystal face, as was done in previous discussion. The diffracted images in question are then designated by the double indices $+1, +1$ and $-1, -1$.

Observations were made, using beams of helium at temperatures of 100° K and 290° K, at various glancing angles of incidence up to 70°. Theory and experiment are again in substantial agreement. But experimental conditions are on the whole less favorable than when the first-described arrangement (that of Fig. 251) is employed. The reason for this is that with the second arrangement either the glancing angle of incidence or the glancing angle of the diffracted beam, or both angles, must be large, over 45°, whereas in the first arrangement small glancing angles could be used and were exclusively used. Imperfection of the crystal surface—that is, roughness—has a more damaging effect toward atom waves at large glancing angles than at small ones, just as it has toward light-waves. Moreover, roughness favors the long wave-length end of the spectrum relatively to the short end. Long waves are more highly reflected or diffracted. It is consequently not sur-

prising that some of the observed maxima should have been found at angles of diffraction which exceed the calculated angles by several degrees.

In a subsequent and concluding investigation Estermann, Frisch, and Stern* succeeded in procuring a homogeneous, or "monochromatic," incident beam, and with this performed diffraction experiments similar to those above described. The homogeneous beam was procured by two different methods in succession:

By one method the original heterogeneous beam from the gun was diffracted by a first crystal and a homogeneous beam was selected out of a portion of one of the diffracted spectra. This beam was then used as the incident beam in the final diffraction experiment involving a second crystal.

By the other method the original heterogeneous beam from the gun was passed through a so-called "velocity analyzer" before it was incident upon the crystal in the final diffraction experiment.

A "velocity analyzer" is an apparatus consisting of two parallel disks mounted upon one and the same axis at a distance of several centimeters from each other. Each disk has a large number of radial slots spaced at equal intervals along its periphery. The whole system is rotated at a high constant speed. An atom or molecule passing through a given slot of the first disk will in general be stopped by a tang at the periphery of the second disk. But if the velocity of the particle falls within a given fairly narrow range, the particle will pass through the next following slot of the second disk. A stream of particles which is approximately homogeneous in velocity thus energes from the analyzer. And this stream constituted the incident beam of the final diffraction experiment.

The velocity analyzer, as an instrument, had been originated a decade earlier by Stern and had in the meantime been employed also by other investigators.†

By way of concluding the present section be it mentioned that R. M. Zabel‡ has carried out experiments involving the diffraction of helium atoms by rocksalt, studying in particular the effect of exposing the crystal to moisture before use. He also experimented somewhat with beams of neon and argon. An ionization gauge was used as a detector.

19. The Diffraction of Hydrogen Atoms.—In the experiments which were discussed in the previous section, the hydrogen which was used was ordinary molecular hydrogen, H_2—the stable gas. T. H. Johnson§ has worked

* *Loc. cit.*

† O. Stern, *Zeit. f. Phys.*, **2**, 49, 1920; Costa, Smyth, and Compton, *Phys. Rev.*, **30**, 347, 1927; J. A. Eldridge, *ibid.*, p. 931; and Ellett, Olson, and Zahl, *ibid.*, **34**, 493, 1929.

‡ *Phys. Rev.*, **42**, 218, 1932.

§ *Jour. Franklin Inst.*, **206**, 301, 1928; **207**, 629 and 639, 1929; **210**, 135, 1930; and *Phys. Rev.*, **37**, 847, 1931.

with atomic hydrogen, H, which is transiently formed when molecular hydrogen dissociates due to collision in an electric discharge. A discharge tube of the Wood type* was used as the "gun"; tubes of this type are commonly some 5 or 10 mm in diameter and about 2 mm in length. They operate at a pressure of a few tenths of a millimeter and are constantly supplied with fresh hydrogen from an electrolytic generator through a capillary connected near one of the electrodes. Atomic hydrogen is copiously produced in the middle portion of the tube. When, as in Johnson's experiments, the tube serves as an atom gun, a slit is provided in the glass wall of the tube. A stream of particles issues from this slit into an inclosure which is rapidly and continuously evacuated. The stream consists mostly of hydrogen atoms but may include other particles, especially molecules and positive ions. The ions are drawn aside by a suitable electric field. A portion of the stream of neutral particles passes through a second slit and then strikes a crystal. The reflected and diffracted beams are received upon a plate of glass upon which a coating of molybdenum trioxide has been previously deposited by smoking the plate over a strip of metallic molybdenum burning in an oxygen-gas flame. Molybdenum trioxide is white. Where a beam of atomic hydrogen strikes the plate, the coating of MoO_3 is reduced to MoO_2, which is blue. Molecular hydrogen does not affect the original coating and hence such molecular hydrogen as may be present does not enter into the final result. A permanent record of the effects produced by the atomic hydrogen may be procured by photographing the oxide-coated plate through a suitable color filter.

H. Kerschbaum† has shown that beams of atomic hydrogen affect Schumann plates and that by using these plates the beams may be photographed directly. Kerschbaum succeeded in obtaining specularly reflected beams but did not develop his method sufficiently to find evidence of diffraction.

Johnson, when using LiF crystals, and having the beam incident normally, obtained the complete pattern of first-order spectra which would be expected from a crossed grating formed by like ions of the surface layer of the crystal. When working at a glancing angle of either 30° or 45°, he obtained diffraction patterns with streamers radiating out from the central image somewhat as shown diagrammatically in our Figure 254. Each streamer represents a diffracted continuous spectrum. Thus the investigations of Johnson, like those of Stern and his collaborators, give conclusive evidence of the wave nature of atoms.

Johnson, through the use of his indirect photographic method, also discovered diffraction effects attributed by him to the secondary or mosaic structure of the LiF crystal.

* R. W. Wood, *Phil. Mag.*, **42**, 729, 1921. † *Ann. d. Phys.*, **2**, 201 and 213, 1929.

20. The Reflection or Diffraction of Metallic Atoms.—A. Ellett and his students, H. F. Olson, H. A. Zahl, and R. R. Hancox, have carried out numerous interesting investigations upon beams of metallic atoms.* We shall next give an outline of these. Be it stated in advance, however, that with metallic atoms, which have relatively large mass and correspondingly small associated wave-length, direct diffraction effects are difficult to detect. However, the occurrence of specular reflection has been well established by the foregoing investigators, although certain apparent deviations from the ordinary law are observed. And specular reflection of atoms by a crystal, as distinguished from diffuse reflection, must be regarded as diffraction of zero order, and constitutes evidence for the wave nature of the atom.

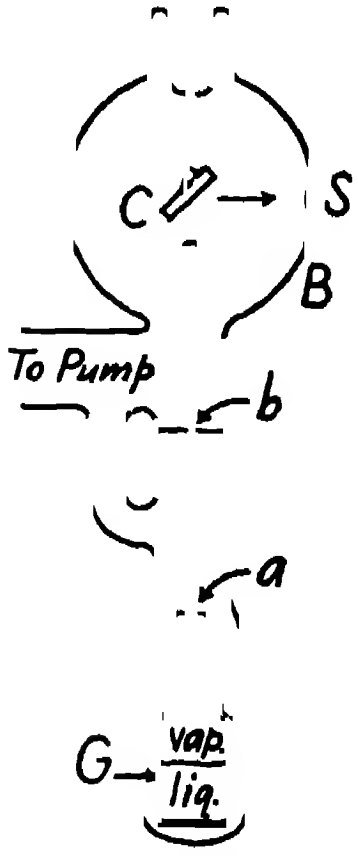


FIG. 257.—Reflection of metal atoms (Ellett and Olson).

Figure 257 shows the apparatus used in the first investigation of the series, by Ellett and Olson,† who experimented chiefly with the reflection of cadmium atoms by rocksalt. The atom gun, G, was a "boiler" in which cadmium was melted and vaporized. A stream of atoms issued from the aperture a, of .5 mm diameter, which was located in a diaphragm at the top of the boiler. The stream then passed through a second aperture, b, of the same diameter, in another diaphragm. The diaphragms were cooled by liquid air to prevent scattering of atoms at the edges of the apertures. The stream of atoms was incident upon the cleavage face of a rocksalt crystal, C. The angle of incidence was varied from trial to trial. The atoms after leaving the crystal struck the inner surface of the glass bulb B, which was surrounded by liquid air. The atoms stuck to the cooled glass surface and formed an opaque, or partially opaque, deposit. The formation of this could be observed from the outside. A slightly modified form of the apparatus permitted more convenient determination of angles. The crystal was heated in vacuum before a run in order to free the surface from adsorbed gas. When the surface was "clean," a considerable portion of the incident beam of atoms was found to be specularly reflected; the angle of reflection equaled the angle of incidence as nearly as they could then determine, and a sharply definied spot .8 mm in diameter was formed at S. Diffuse scattering also occurred, as revealed by a

* Ellett and Olson, *Phys. Rev.*, **31**, 643, 1928; Ellett, Olson, and Zahl, *ibid.*, **34**, 493, 1929; H. A. Zahl, *ibid.*, **36**, 893, 1930; Zahl and Ellett, *ibid.*, **38**, 977, 1931; R. R. Hancox, *ibid.*, **42**, 864, 1932.

† *Loc. cit.*

light deposit extending over all of that half of the bulb which faced the bombarded surface of the crystal.

Experiments with mercury placed in the boiler showed that specular reflection of mercury atoms also occurs. With sodium, on the other hand, only diffuse reflection was found. This is probably because sodium atoms are adsorbed by the surface and subsequently re-emitted, and is in line with the findings of J. B. Taylor to be mentioned at the end of the section.

In a continuation of the reflection experiments with cadmium, Ellett, Olson, and Zahl* found that when the original beam was incident at 45°, about 17 per cent of it went into the small specular spot. To ascertain the density of the deposits the deposits were run through a densitometer or were photographed.

In one set of experiments a velocity analyzer was introduced to ascertain the distribution in velocity of the atoms of the specular beam. The investigators concluded from their observations, at the time, that in the process of reflection there were selected out of the original incident beam, for specular reflection, only atoms having velocities falling within a given range. This conclusion appeared, moreover, to be borne out by the fact that in certain other, double-reflection experiments, which they likewise performed, a specular beam from the first crystal appeared to be specularly reflected by the second crystal only when the angle of incidence was the same upon both crystals. A later investigation by H. A. Zahl,† who worked in addition with beams of zinc atoms, however, led to the conclusion that the various phenomena which are involved are more complicated than had at first been supposed.

In subsequent investigations Ellett and Zahl, and R. R. Hancox,‡ have resorted to an ionization gauge for measuring intensities of the reflected beams, in place of the deposit method, and have reflected beams of mercury and cadmium, from crystals of NaCl, KCl, KBr, KI, LiF, LiCl, and NaF, varying the temperature both of the beam and of the crystal, from run to run. They find that the reflected beam usually lies several degrees nearer the normal to the crystal surface than the incident beam. Their hypotheses as to the possible explanations of their findings we cannot here enter into.

J. B. Taylor§ has investigated the reflection of the alkali atoms lithium, potassium, and caesium from crystals of sodium chloride and lithium fluoride. Alkali atoms are subject to ready adsorption, therefore, in order to avoid, if possible, momentary adsorption of the atoms upon the crystal surface, the crystals were held at various temperatures up to 500° C. No trace of specular reflection or diffraction was found, however. The number of

* *Loc. cit.*

† *Loc. cit.*

‡ *Loc. cit.*

§ *Phys. Rev.*, **35**, 375, 1930.

atoms leaving the crystal surface in any direction was found to follow the law of Knudsen, that is, the number is proportional to the cosine of the angle between the normal to the surface and the given direction, to a very high degree of accuracy, .1 per cent. It is probable that momentary adsorption accounts for the failure of the alkali atoms to be specularly reflected. For measuring the intensity of the beams, an electrical method was employed. The method employed is one which is based on an observation of Langmuir and Kingdon's and is uniquely applicable to streams of alkali metal atoms. The atoms after reflection enter a hollow metal cylinder through a slot and strike an incandescent tungsten filament within the cylinder. Each atom which strikes the filament gives up an electron. The residual positive ion, impelled by a low electric field which is applied between the filament and the cylinder, then travels to the cylinder. An ion current results and the strength of this current is measured.

21. Experiments with Positive Ions.—We shall speak first of the experiments of A. J. Dempster* upon the reflection or diffraction of hydrogen ions:

When a discharge tube has a perforated cathode, rays of positive ions, which are known as "canal rays," appear behind the cathode. When, as in Dempster's experiments, a discharge tube is filled with hydrogen, the positive rays which are formed include hydrogen atom ions, or protons, and hydrogen molecule ions, and also triatomic hydrogen ions. Dempster allowed a stream of hydrogen canal rays to fall at nearly grazing incidence upon single crystals, especially calcite. A maximum potential of 40,000 volts was applied to the tube, and there were ions present in the stream ranging from 40,000 volts down to 15,000 volts and lower. The charging-up of the crystal which would naturally be expected to occur when, as in these experiments, a target is used which is an insulator was supposedly obviated by placing a wire over the crystal and in contact with it except in a small region where the beam was incident. Secondary electron emission from the wire was depended upon to dissipate the charge. A photographic plate was placed in the path of the beams leaving the crystal, and the photographs thus obtained showed streamers which were interpreted as indicating that some type of diffraction was taking place. It was at first believed that the photographic effects arose entirely from the protons which were present in the streams leaving the crystal, but later a positive-ray analysis showed that other positive ions present also produced effects. The results did not permit of drawing definite conclusions of any sort.

Y. Sugiura† also undertook to diffract protons. He obtained his stream as follows: Hydrogen from an electrolytic generator was allowed to diffuse

* *Phys. Rev.*, **34**, 1493, 1929, and **35**, 1405, 1930; *Nature*, **125**, 51 and 741, 1930.

† *Sci. Papers of the Inst. for Phys. and Chem. Res.* (Tokyo), **16**, 29, 1931.

through the heated end-wall of a small tube made of palladium. Hydrogen passes through hot palladium rather freely. A ring-shaped filament situated beyond the end of the palladium tube furnished electrons which by bombardment accomplished the double purpose of heating the end-wall of the palladium tube and of ionizing the hydrogen after it had passed through this end-wall. The ionizing process yielded mostly H^+ ions, but Sugiura found that there were also some H_2^+ ions present. Both types of ions were accelerated by an electric field, and after passing down a collimating tube were separated from each other by the well-known magnetic deflection method. The H^+ ions then passed through two collimating slits and thereupon fell at grazing incidence upon a polycrystalline deposit of platinum or of tungsten upon glass. The beams which emerged from this surface in nearly grazing directions were then examined by passing across them a receptor consisting of a Faraday box, connected to a Compton electrometer. Rather faint maxima superposed upon a broad background were recorded.

Sugiura gave an interpretation to his results to which exception must be taken. The incident beam is homogeneous and the platinum or tungsten deposit is polycrystalline with the microcrystals presumably oriented at random. Sugiura believed that he hence had the conditions of Thomson's method of electron diffraction except that instead of having the rays pass through the metal film placed perpendicularly he had the rays grazing the film and then emerging from it on the same side on which they were incident. That is, he supposed that Bragg reflection took place in each microcrystal. The accelerating potentials which he used were low, however, ranging from 290 to 450 volts. At these voltages the penetrating power of protons is probably too low to yield space-lattice diffraction effects.

The effects must be rather those arising from diffraction by the surface lattices of the microcrystals. Consider a monochromatic beam grazing a plane grating and consider the direction of the first-order spectrum. If the grating is rotated, the first-order spectrum moves to larger angles of deviation. Thus when there are minute surface lattices effective having all possible orientations, there will be discontinuities in the intensity plotted as a function of angle. It is probably these discontinuities which Sugiura observed. The case has been treated mathematically by M. von Laue.*

Subsequently, E. Rupp,† and then Meibom and Rupp,‡ successfully applied Thomson's method of electron diffraction in its original form to the diffraction of protons. They used accelerating potentials ranging from 180 to 250 kilovolts and caused the protons to pass through a very thin gold film

* *Zeit. f. Kristallog.*, **82**, 127, 1932; a preliminary discussion was given in *Die Naturwiss.*, **19**, 951, 1931.

† *Zeit. f. Phys.*, **78**, 722, 1932. ‡ *Annalen*, **17**, 221, 1933.

placed perpendicularly to the beam. At a distance of 3.5 m beyond the film they erected a photographic plate and obtained a system of small rings, having diameters up to about 5 mm. They found agreement with the De Broglie relation to within 2 or 3 per cent. The De Broglie wave-length of 200 kilovolt protons is only 6.4×10^{-12} cm! Thin films of mica, in place of gold, were also tried. Rupp, in his first work, also undertook to make atom-form factor determinations for protons from his photographs.

G. E. Read,* R. W. Gurney,† R. B. Sawyer,‡ and A. Longacre § have performed reflection experiments with the positive ions of several metals. A suitable compound containing the metal was placed as a coating upon a wire and the compound was then brought to incandescence by heating the wire electrically. Under these conditions positive ions of the metal are given off. The ions were accelerated by applying a potential difference between the wire and the wall of the gun. Beams of positive ions so obtained were reflected or scattered from platinum foil used as a target. Platinum foil is polycrystalline with the microcrystals arranged at random. Lithium ions have also been reflected in this manner from a deposit of nickel evaporated on to a tungsten foil. Such a deposit is polycrystalline, but in the case of this evaporated deposit the microcrystals are presumably arranged with their (1 1 1) planes parallel to the tungsten backing foil.

The streams of positive ions which emerge from the target are received into a Faraday box or simply upon a flat, not very clean, metal surface, to which they stick. The receptor is connected to a galvanometer or electrometer.

Some of the scattering curves which have been recorded show marked lobes. The reason for the occurrence of these lobes is not known.

* *Phys. Rev.*, **31**, 629, 1928.

† *Ibid.*, **32**, 467, 1928.

‡ *Ibid.*, **35**, 1090, 1930.

§ *Ibid.*, **42**, 906, 1933.

APPENDIX A

THE METHOD OF LUNES AND THE CORNU SPIRAL

(Continuation of Chap. 3)

We shall begin by explaining in detail why it is permissible to regard the disturbance from each elementary lune as originating from the corresponding minute element of the horizontal arc drawn across the wave front.

As previously stated, it is never necessary to make use of more than, let us say, twenty convolutions of the vibration curve. The first twenty convolutions are contributed by the first forty half-period lunes and these all lie close to the pole of the wave. Looking from the field point toward the wave front, let M_0 (Fig. 258) be the pole of the wave and let F mark the

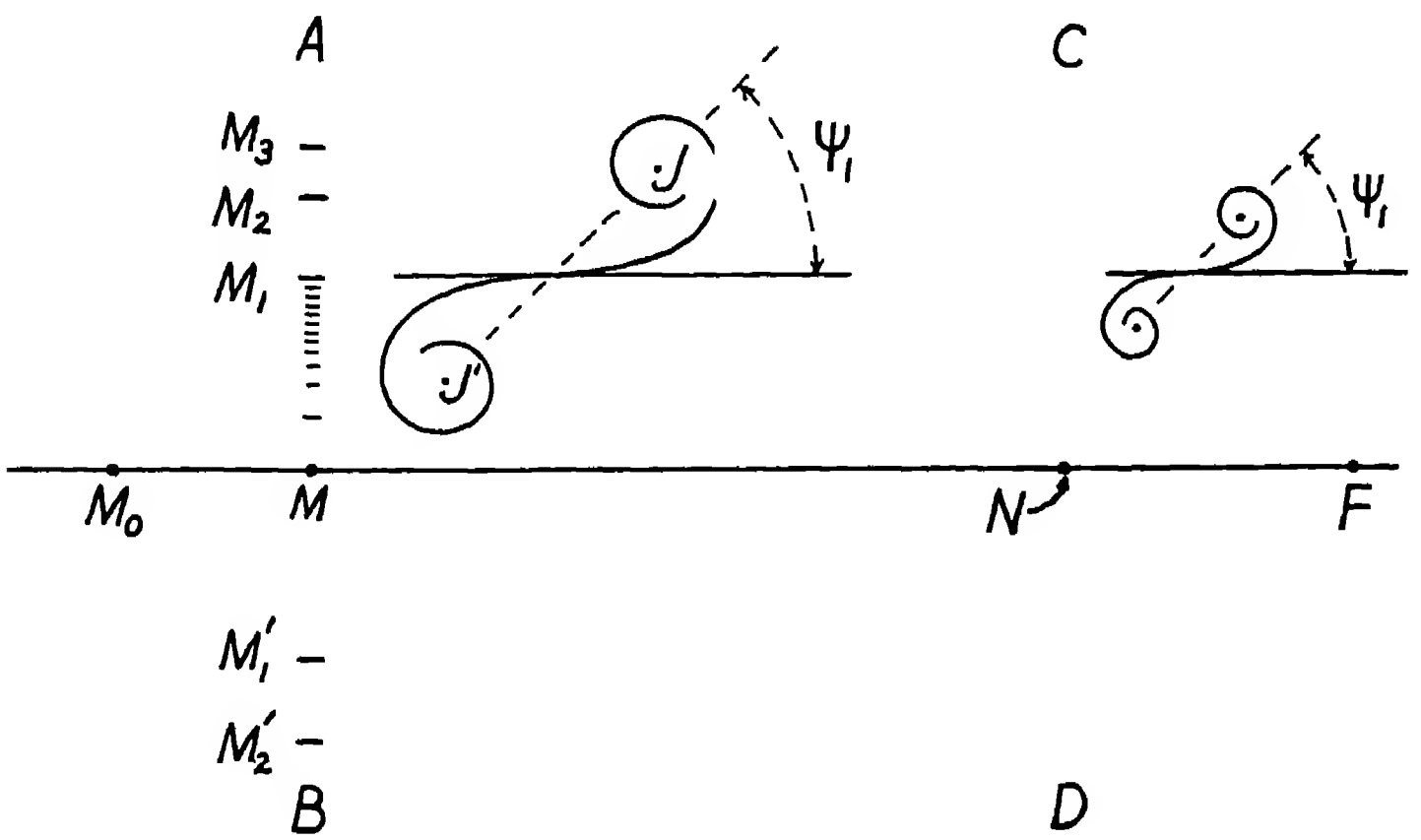


FIG. 258

position of the outer edge of the fortieth half-period lune. Let AB and CD be two *elementary* lunes which both fall between the pole of the wave and F. (These elementary lunes are of necessity represented in greatly exaggerated proportion.)

Beginning at M on the central horizontal arc, proceed vertically along the elementary lune AB, dividing this into half-period sections as indicated by $M_1, M_2 \ldots . M_1', M_2'$, etc., and then subdivide these sections into minute sections as indicated. With these minute sections as a basis, construct a vibration polygon for this elementary lune, and then, passing to the limit, construct a vibration curve. The entire curve will be minute because the

width of the lune is minute. Let a disturbance originating from M be represented by a vector drawn horizontally to the right. Then the vibration curve will evidently have the general form of the curve shown to the right of AB. The closing vector drawn from the lower asymptotic point to the upper one represents the resultant disturbance arising from the entire elementary lune.

Now, in the same way, construct a vibration curve for the elementary lune CD. Let a disturbance originating from N be represented by a vector drawn horizontally to the right. Then the vibration curve will have the general form shown by the curve to the right of CD. The closing vector joining the two asymptotic points of this curve represents the resultant disturbance from the elementary lune CD.

The second curve is smaller than the first in the proportion in which the width of CD is less than that of AB.

Moreover, the curves will be practically similar in form, and hence the angle indicated by ψ_1 in each curve, which the closing vector drawn between the asymptotic points makes with the horizontal direction, will be practically the same for both curves. Hence when we consider the resultant disturbance from each elementary lune as originating from the corresponding element of the horizontal arc drawn across the wave front, we will be omitting throughout a constant phase angle ψ_1, and the omission of this constant angle will not affect the final result. If the angle were taken account of, the vibration diagram for the complete wave front would be simply rotated through an angle ψ_1 with reference to the orientation which we have previously considered the curve to have (see Fig. 259).

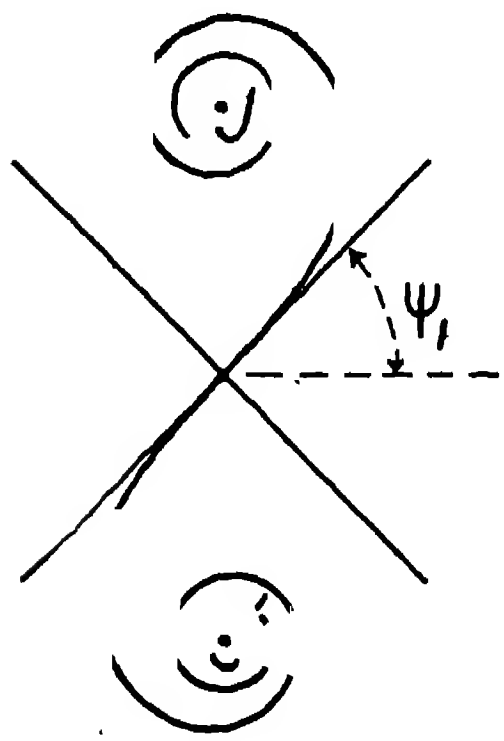


FIG. 259

The vibration curves for the elementary lunes will each be practically Cornu spirals, and hence the angle ψ_1 will have practically the value $\pi/4$, this being the angle which the vector joining the asymptotic points of the Cornu spiral makes with the horizontal.

Let us now evaluate the previously introduced "equivalent height," h, of an elementary lune: Consider the vibration curve for an elementary lune to be a minute Cornu spiral. The vector joining the asymptotic points of this spiral has a half-length equal to $\frac{1}{2}\sqrt{2}$. About the origin of co-ordinates as center and with a radius of length $\frac{1}{2}\sqrt{2}$, extending to, let us say, the *upper* asymptotic point, describe an arc until this intersects the first semi-convolution of the spiral. The intersection will occur at a point which measured along the curve lies a little beyond the point $v=\frac{1}{2}\sqrt{2}=.707$,

whereas the first semiconvolution ends at $v=\sqrt{2}=1.41$, and to this length of arc 1.41 corresponds the first half-period section of the lune measured vertically. Therefore a section of the lune extending just a little beyond *half* of the first half-period section above the central horizontal arc and an equal distance below the central horizontal arc will contribute an effect of the same magnitude as the effect contributed by the entire lune.

Fixing our attention on the first half-period section of each elementary lune near the pole, this section will have practically the same height throughout. Therefore we may consider that the effect of the entire wave may be replaced by a horizontal strip of uniform height taken across the wave front. If we proceed far from the pole horizontally, the half-period sections of the elementary lunes increase slightly but appreciably in length, because, as the radius vector r from the field point to the horizontal arc increases in length, we must go a greater distance vertically before attaining the length $r+(\lambda/2)$. Hence it is sometimes alternatively stated that the horizontal strip should be thought of as widening slightly in passing from the pole of the wave outward. There is, however, no real distinction to be drawn between the two points of view thus represented. It is a basic feature of the theory upon which the use of the Fresnel integrals and the Cornu spiral rests that we do not need to concern ourselves in all detail with what happens far from the pole of the wave, and it is immaterial whether we consider the horizontal strip to be uniform or to widen slightly far from the pole.

The scheme of considering the wave front by lunes was originated by Fresnel.*

When the incident wave front is plane or cylindrical, each lune becomes a vertical strip. The term "laminar zones" has been introduced by A. Schuster† to denote such strips.

The Fresnel integrals and the Cornu spiral apply also to plane waves and to cylindrical waves. Furthermore, it is sometimes erroneously stated that they apply only to plane or cylindrical waves. The additional statement sometimes made in the same connection that when a slit is used as a source this gives rise to cylindrical waves is also false, or at least misleading. When a slit is illuminated in the usual manner, each point of the slit is an independent source of spherical waves.

We shall conclude by deducing from a somewhat different point of view than the one taken previously that equation (7) of chapter 3, when used as a basis for constructing a vibration curve, leads to a Cornu spiral.

* *Œuvres compl.*, **1**, 196. Fresnel used the word *fuseau*, which means a "spindle," i.e., something which is thick in the middle and tapering toward both ends.

† See, e.g., Schuster and Nicholson, *Introduction to the Theory of Optics* (3d ed.; Arnold & Co.), pp. 92 ff.

The amplitude term of equation (7), when taking account of equation (8), may be written:

$$(1) \qquad \frac{k\,A\,h\,dq}{ab} = \frac{k\,A\,h\,dv}{abC} = E\,dv\,,$$

where E is a constant introduced by way of abbreviation. In the phase term of equation (7) write for $2\pi\delta/\lambda$ its equal, namely, $\pi v^2/2$, as given by equation (11). Equation (7) thus becomes:

$$(2) \qquad ds = E \cdot dv \cdot \cos\left[2\pi\left(\frac{t}{\tau} - \frac{a+b}{\lambda}\right) - \frac{\pi v^2}{2}\right].$$

Here $E\,dv$ is the amplitude of an elementary simple harmonic motion which is to be laid off as a vector of length dv. Applying the formula for the cosine of the difference of two angles, the foregoing equation becomes:

$$(3)\; ds = \left[E\cos\frac{\pi v^2}{2}\,dv\right]\cos 2\pi\left(\frac{t}{\tau} - \frac{a+b}{\lambda}\right) + \left[E\sin\frac{\pi v^2}{2}\,dv\right]\sin 2\pi\left(\frac{t}{\tau} - \frac{a+b}{\lambda}\right).$$

Each of the two terms on the right is a simple harmonic motion. The square brackets are the amplitudes of the two simple harmonic motions. The phases of the two motions differ by $\pi/2$ because of the occurrence of respectively the cosine and sine of the same function of the time, t. The sum is a vector sum. The trigonometric factors which contain the time contain this as the only variable. In particular, they do not contain any variable depending on distance measured across the wave front. Hence in integrating across the effective portion of the wave front we may take these factors outside of the integral sign. Accordingly, integrating across the wave front, the resultant disturbance becomes:

$$(4)\; s = \left[E\int_{v_1}^{v_2}\cos\frac{\pi v^2}{2}\,dv\right]\cos 2\pi\left(\frac{t}{\tau} - \frac{a+b}{\lambda}\right) + \left[E\int_{v_1}^{v_2}\sin\frac{\pi v^2}{2}\,dv\right]\sin 2\pi\left(\frac{t}{\tau} - \frac{a+b}{\lambda}\right).$$

Each term of this equation is again one of two simple harmonic motions which differ in phase by $\pi/2$. Each term contains one of the Fresnel integrals and when these are plotted at right angles to each other they yield the Cornu spiral.

APPENDIX B

THE PATTERN DUE TO A SINGLE STRAIGHT EDGE; LABORATORY DIRECTIONS

(Submitted in Connection with Chap. 4, Sec. 1)

The directions given below are, except for necessary minor changes, those which the author has for several years given his students. Following the directions proper, supplementary remarks will be made, these relating especially to alternative methods of carrying out the experiment.

Directions

This experiment falls into three parts:

I. Determination of the wave-length of the light which is to be used in the experiment proper.

II. Photography of the pattern due to a single straight edge, and measurement of the photographic plate.

III. Comparison of the results of experiment with theory, by applying the Cornu spiral.

PART I

Obtain a coarse transmission grating—for example, a replica having about 1000 lines per centimeter—and determine the grating constant of this under a traveling microscope or under a microscope provided with an eyepiece micrometer. Then employ the grating for determining the wave-length of the source of light in the manner about to be described.

Referring to Figure 260, S is the primary source of light—an incandescent lamp with a spiral filament—C is a condensing lens, and O a slit, perpendicular to the plane of the paper, or vertical. Focus the image of the lamp filament upon the slit.

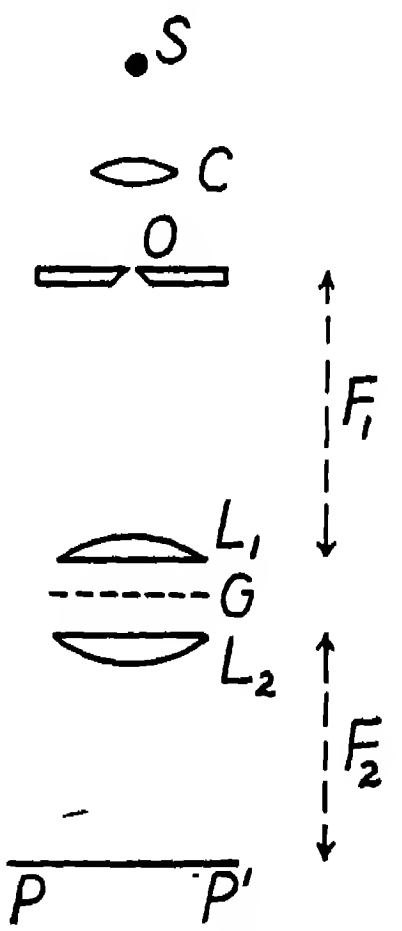


Fig. 260

Making use of a mounting furnished you, place the grating (G in the figure) between the two lenses L_1 and L_2, adjusting the distance indicated by F_1 to be equal to that of the focal length of the lens L_1. At the distance F_2 equal to the focal length of the lens L_2 erect the diffusing screen, PP', which has been furnished (a piece of white paper pasted onto an old photographic plate, the whole fitting into a plate-holder for which a mounting is provided at PP', this plate-holder being the same which will later hold the

photographic plate). Your apparatus now constitutes a spectrograph, though, to be sure, a somewhat crude one. The central image should be plainly visible at the center of the diffusing screen (or photographic plate), and the first- and second-order continuous spectra should be visible to the right and left. Try the focus and set at the best position.

Insert the Noviol (no violet) glass filter furnished you, somewhere between the primary source S and the slit O, and observe the effect which the insertion of this filter has upon the spectra seen upon the diffusing screen (the filter cuts out all of the violet and a part of the blue). The photographic plates which you have, ordinary ones, not panchromatic or orthochromatic or the like, are not sensitive to the green, yellow, or red, but are sensitive to the remaining portion of the blue—and this falls into a sufficiently narrow spectral region so that for the purpose of the present experiment, which does not require highly monochromatic light, the light which is photographically effective may be considered monochromatic. Insert a photographic plate in the plate-holder and expose for several seconds, and then develop, fix, and dry the plate.

Measure upon the plate the distance from the central image to each spectral image. This distance, divided by the distance F_2, gives the tangent of the angle of diffraction. Having deduced the tangent, look up the sine in tables, and obtain for each spectral image a value of the wave-length by applying the grating law, which, in the present case, takes the simple form $m\lambda = a \sin \theta$. Why? Average the values so obtained. Finally, make a rough estimate of how much your measurements may be in error and by how much in consequence your wave-length determination may be in error.

PART II

Use the same source of light, condensing lens, and slit as before, and have the Noviol filter again in place. Referring to Figure 261, erect vertically at the position B the small bar furnished you, which has a cross-section in the form of a keystone. Place the sheet of metal, M, which is supported by a rod from the wall, so that this sheet extends in front of the bar about half-way, as indicated in the figure; but the sheet must not touch the bar. The right-hand edge of the bar now acts as a single straight edge.

Eventually the position of the edge of the geometrical shadow must be ascertained. Steps must therefore be taken now which will make it possible to locate this position in the pattern later. It is the position where a straight line drawn from the slit, O, touching the free edge of the bar, B, strikes the photographic plate. This point is represented by E in the figure. By sliding the sheet of metal M on the slide arranged for the purpose, over to a corresponding position M', the left-hand edge of the bar can be made to

serve as a single straight edge and E' will then be the edge of the geometrical shadow. Calling W the width of the bar, we have:

$$EE' = W\,\frac{a+b}{a}, \tag{1}$$

where a and b are the distances indicated. By making one exposure when the sheet of metal is in the position M, and another when it is in the position M', the edge of the geometrical shadow can eventually be located in each of the two diffraction patterns which are to be photographed—in a manner which will appear as we proceed.

Measure the distances a (*ca.* 120 cm) and b (*ca.* 140 cm) with a meter stick, and measure the width of the bar (*ca.* 5 mm) carefully upon a traveling microscope or comparator.

Expose for about ten minutes with the metal sheet in the M position. Then cut off the light by interposing some opaque object at some convenient point in the beam, shift the metal sheet to the M' position, and expose again for an equal length of time. Having begun the first exposure, avoid even touching the bar or the plate-holder, for fear of displacing them, until the second exposure has been completed. Develop, fix, and dry the photographic plate.

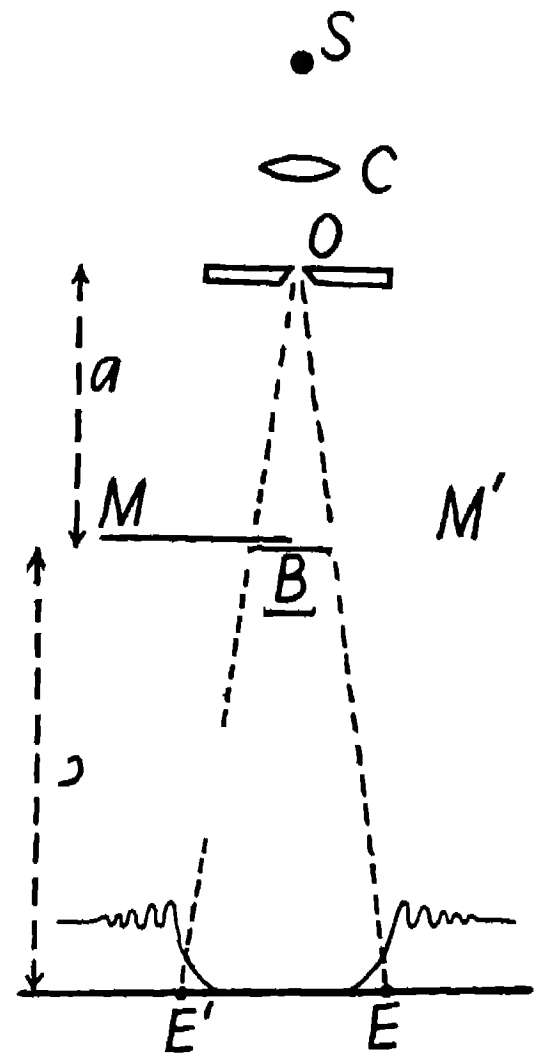


Fig. 261

The two diffraction patterns, each for a single straight edge, which your photograph has upon it, are located with reference to each other as illustrated in Figure 262. The region between the two patterns is a region of shadow for both patterns. This region is, of course, light upon the photograph, since this is a negative. Further, the region to the right of the right-hand pattern is a region of uniform illumination for the right-hand pattern, and the region to the left of the left-hand pattern is a region of uniform illumination for the left-hand pattern. Place the photographic plate upon the comparator, orient it properly, and focus the microscope until the parallax has been removed. Now turn the comparator screw until the field of vision of the microscope falls into the uniformly illuminated region to the right of the right-hand pattern. You are now ready to begin making measurements. Turn the comparator screw so that the field of vision moves toward the left, and, as you proceed, make several settings upon each maximum and each minimum of the right-hand pattern which is distinct enough to

make settings upon. Having completed your measurements in the right-hand pattern, continue, through the middle region of shadow, and make settings in the left-hand pattern, measuring straight across the entire plate from right to left. When you have finished, return to the right-hand pattern, and, by way of check, to make sure that nothing has been disturbed during the course of your measurements, make several settings upon some one maximum or minimum which permits of having accurate settings made upon it—for example, the second maximum.

Tabulate your data as shown in Table IX. Having completed column 4, which gives the distance between the two maxima of the same number or

TABLE IX

1	2	3	4	5
Designation of Fringe	Individual Settings	Mean	Difference, Left *minus* Right	Distance from Edge of Geometrical Shadow, i.e., Column 4 *minus* EE', and then divide the difference by 2
5th max. { left. . . . . right				
4th min. { left . . . . . right.				
4th max. { left. . . . . right.				
Etc.				

the two minima of the same number, in the two patterns, subtract from column 4 the distance EE' as calculated by applying equation (1). The difference obtained is obviously double the distance of the given maximum, or given minimum, from the edge of the geometrical shadow. Therefore divide this difference by 2, thus obtaining the distance itself, and enter the various distances in column 5. Lay off the distances, as given by column 5, to as large a scale as convenient upon a sheet of co-ordinate paper, drawing a figure which resembles one-half of Figure 262, let us say the right-hand half, that is, placing the edge of the geometrical shadow, E, at the left in your figure.

You have now completed your experimental work and recorded the result.

PART III

Obtain a graph of the Cornu spiral to a reasonably large scale (a blueprint is usually furnished). You are to measure with a meter stick the

length of the closing vector drawn from certain points of the spiral to the asymptotic point J in the upper right-hand quadrant of the graph. Begin with the point marked 3 in the upper branch of the spiral, measure to J, and record the length; then from 2.5 to J; then from 2 to J; etc., proceeding by intervals of .5 until you reach the point 1 in the lower branch of the spiral. Thereafter proceed by intervals of .2 until you reach the point 3 in the lower branch, and thereafter proceed by intervals of .1 until you reach the point 4. Tabulate your data in two columns: column 1, giving the point of origin of the closing vector, that is, the number on the spiral, or

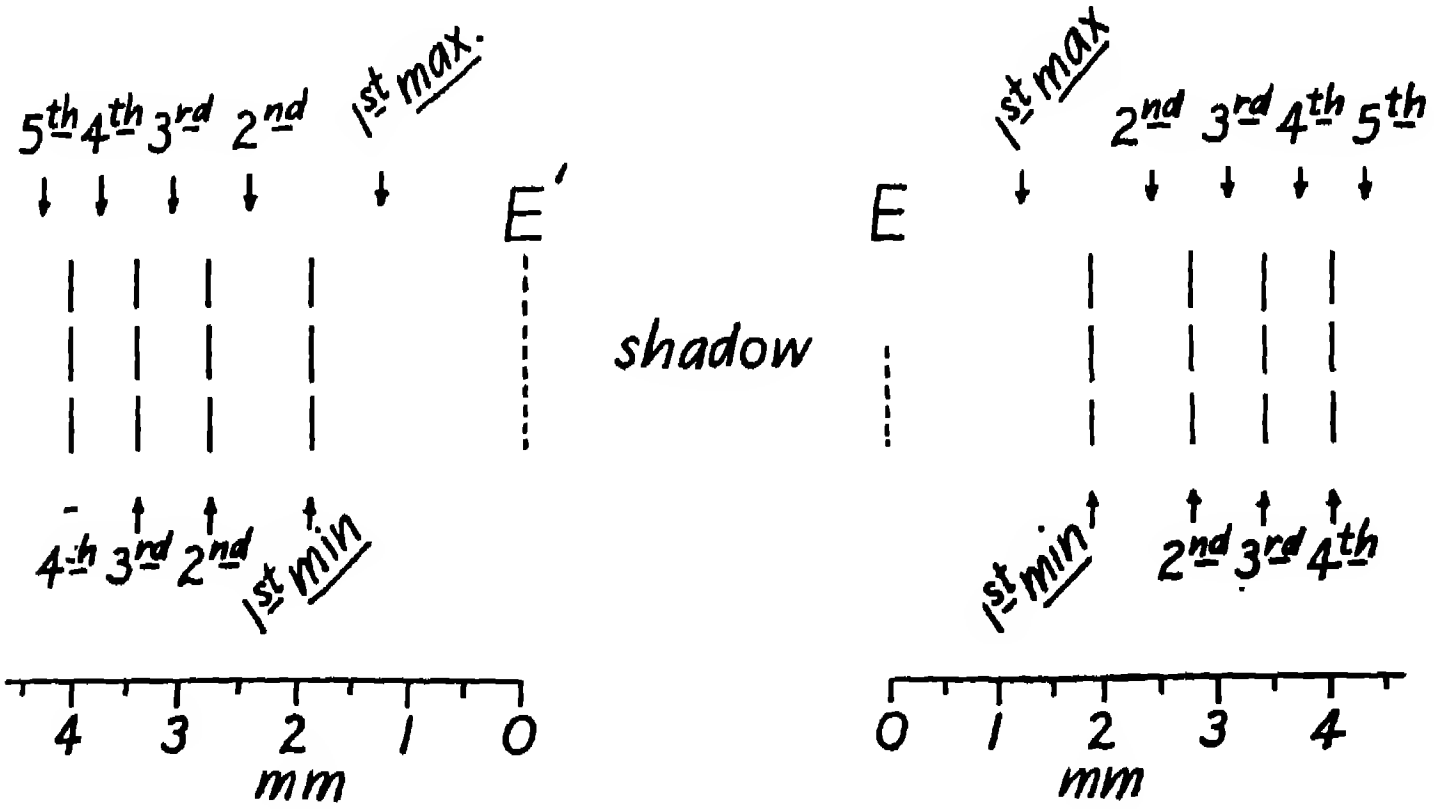


FIG. 262

v-value; and column 2, giving the length of the corresponding closing vector to the point J, representing amplitude of illumination.

Plot your data on co-ordinate paper, column 1 as abscissae and column 2 as ordinates, but omit the first five values, noting, however, that these increase successively; the exact shape of the curve over this portion is not of interest and by omitting these points the remainder of the curve may be plotted to a larger scale. Begin on the left of your sheet of co-ordinate paper. You should now obtain a curve like that given in chapter 4, Figure 52, curve A. Mark the position on your curve which corresponds to the edge of the geometrical shadow. Further, mark on your curve the abscissa, v-value, at which each maximum and each minimum occurs.

The Cornu spiral is designed to be of general utility, that is, its use is not limited to a set-up of particular dimensions or to light of a particular wavelength. The units of the spiral are v-units, which are proportional to distances measured across the wave front, q-distances, as shown by equation (2) of chapter 4, which is:

$$v = q\sqrt{\frac{2(a+b)}{ab\lambda}}. \tag{2}$$

The factor of proportionality, the radical, is of the dimension length^{-1}. Moreover, distances measured across the diffraction pattern, z-distances, are proportional to q-distances; we have the relation based upon simple geometrical considerations which is given by equation (1) of chapter 4, namely, $q=za/(a+b)$. Eliminating q, we have equation (3) of chapter 4, namely:

$$z=v\sqrt{\frac{(a+b)b\lambda}{2a}}\,. \tag{3}$$

By means of this equation, substituting the values of a, b, and λ which obtain in your experimental work, convert the abscissa, v-value, of each maximum and of each minimum of your theoretical curve into a corresponding z-value, expressed in millimeters, then draw a figure from your theoretical data which corresponds to the figure which you previously drew upon the basis of your observational data (like the right-hand half of Fig. 262). Draw the new figure to the same scale as the old one. Then place one drawing below the other, making the two coincide at one arbitrarily chosen point, let us say at the second maximum. Paste the two drawings together in this position. The other maxima and minima should also very nearly coincide with each other as should likewise the position of the edge of the geometrical shadow as marked upon the two drawings. If the agreement is satisfactory, your work is now complete.

Supplementary Remarks

A mercury arc provided with a filter which passes only the Hg blue lines, or only the Hg violet lines, yields more highly monochromatic light than the incandescent source described above, and such an arc should be used if, or when, it is desired to photograph more than the first five or six fringes. However, with a mercury arc the required exposure time is greater—even with the usual type of quartz mercury arc. (The author has not tried the capillary type of arc.)

The question of exposure time is one of importance in the experiment, since the inexperience of those who commonly perform it often leads to failure and consequent necessity of repetition.

The use of a carbon arc in place of the incandescent lamp would no doubt further shorten the exposure time considerably. The small carbon arcs furnished by microscope-makers for dark-field work burn steadily for a relatively long time, and one of these would probably serve very well.

The work is carried out photographically, partly for the purpose of imparting experience in photographic technique. Also, when the experiment is carried out visually each pair of workers ties up the apparatus for a longer

period of time. And since, in working visually, manipulation of the apparatus during the course of taking observations is necessary, all parts of the apparatus require to be very solidly mounted. In working photographically, on the other hand, the apparatus proper need not be touched at all, and hence great solidity of mounting is not required.

Instead of making two exposures in succession (corresponding to the positions M and M' of the metal sheet), one may alternatively arrange an observation to demonstrate the well-known fact that the external fringes which appear simultaneously upon both sides of a bar are, even when the bar is quite narrow, for all intents and purposes identical with the fringes due to a single straight edge. And having demonstrated this, one is justified in photographing the two sets of fringes *simultaneously*. But in doing this, while one claims to be determining the positions of the fringes due to a single straight edge, one actually works with a narrow bar.

APPENDIX C

THE FRAUNHOFER PATTERN DUE TO A CIRCULAR APERTURE (AIRY)

(Continuation of Chap. 5, Sec. 3)

Referring to Figure 263, suppose MN' is the portion of the actual wave front which passes through the circular aperture, and P is a field point at which we wish to determine the illumination. It will be convenient, for reasons of symmetry, to choose the diffracted wave front as passing through the center of the aperture. Accordingly, consider DE' as the diffracted wave

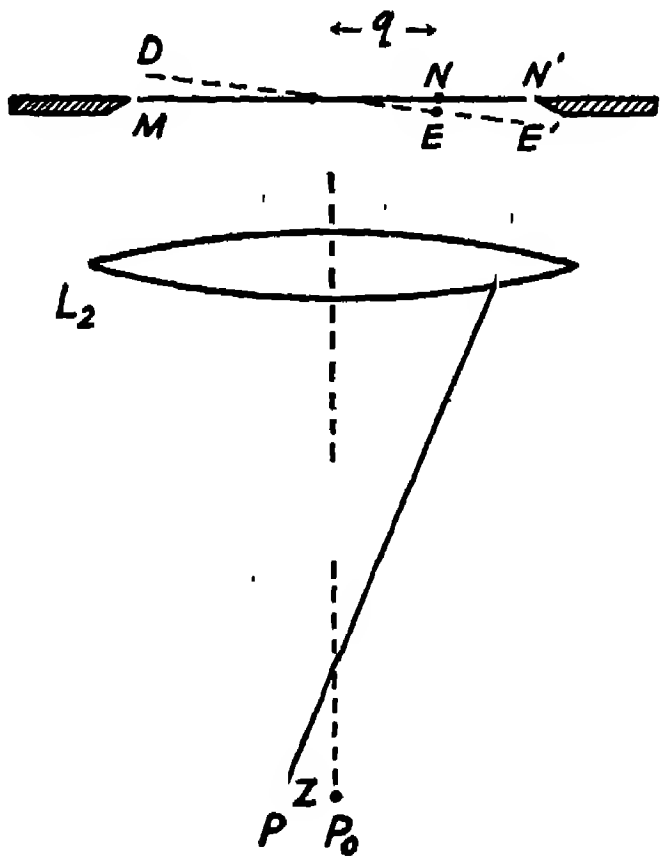


FIG. 263

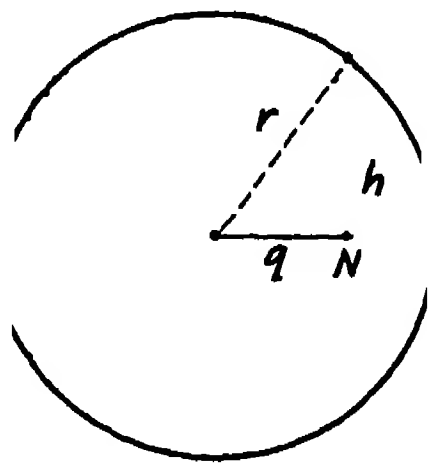


FIG. 264

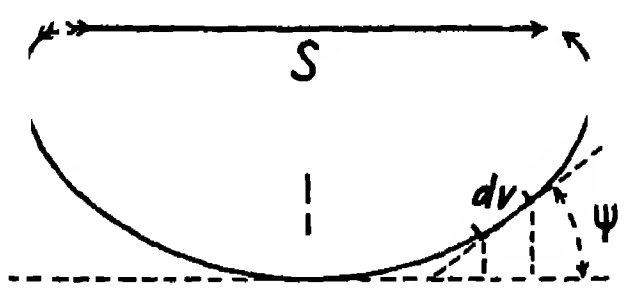


FIG. 265

front with reference to the field point P. Denote the radius of the aperture by r, and the path increment $N'E'$ by Δ, this being the path increment at the edge of the aperture taken with reference to the center. The path increment $\delta = NE$ which obtains at the point N located at a distance q from the center of the aperture is $\delta = (q/r)\Delta = q(z/f)$, where z denotes the distance PP_0 and f denotes the focal length of the lens L_2.

If we consider the actual wave front to be divided into vertical strips of equal width (see Fig. 264), we may construct a vibration polygon, or, passing to the limit, we may construct a vibration curve. This vibration curve will have some such form as shown in Figure 265, and the closing vector S of this curve represents the amplitude of illumination which obtains

at the field point P. Denote the length of an infinitesimal element of arc of this curve by dv and the angle which this element makes with the horizontal by ψ. The angle ψ is the retardation in phase of a disturbance traveling by way of a vertical strip located at N with reference to a disturbance traveling by way of a vertical strip through the center of the aperture. We have $\psi = 2\pi\delta/\lambda = (2\pi\Delta/\lambda)\cdot(q/r)$. The horizontal projection of the element of arc dv is $dv \cos\psi$, and the integral of $dv \cos\psi$ extended over the vibration curve equals the closing vector S, that is:

$$S = \int \cos\psi \cdot dv\,. \tag{1}$$

The element dv is contributed by a vertical strip taken on the wave front and the length of dv is proportional to the area of the strip. Denote the width of the strip by dq and the half-height by h. We are interested only in relative values and hence we may write $dv = h\,dq$. Since $h = \sqrt{r^2 - q^2}$ and $\psi = (2\pi\Delta/\lambda)\cdot(q/r)$, as above noted, equation (1) becomes:

$$S = \int_{-r}^{+r} \left\{ \sqrt{r^2 - q^2} \cdot \cos\left(\frac{2\pi\Delta}{\lambda} \cdot \frac{q}{r}\right) \right\} dq\,. \tag{2}$$

This integral may be expressed as an infinite series. The details of the operation will be given later. The result is:

$$S = \int_{-r}^{+r} \left\{ \qquad \right\} dq = 2\int_{0}^{r} \left\{ \sqrt{r^2 - q^2} \cos\left(\frac{2\pi\Delta}{\lambda} \cdot \frac{q}{r}\right) \right\} dq \tag{3}$$

$$= \pi r^2 \left[1 - \frac{\left(\frac{2\pi\Delta}{\lambda}\right)^2}{2^2 \cdot 4} + \frac{\left(\frac{2\pi\Delta}{\lambda}\right)^4}{2^2 \cdot 4^2 \cdot 6} - \frac{\left(\frac{2\pi\Delta}{\lambda}\right)^6}{2^2 \cdot 4^2 \cdot 6^2 \cdot 8} \cdots \right] \tag{3'}$$

By assigning a succession of values to Δ in equation (3′), or, better, assigning a succession of values to Δ/λ—let us say $\Delta/\lambda = 0, .1, .2$, etc.—we may tabulate the corresponding values of S, the amplitude of illumination, or of S^2, the intensity. (The series [3′] is convergent for all finite values of Δ or Δ/λ.) The tabulated values when plotted yield respectively the curves for amplitude and intensity shown in Figure 86 of chapter 5. The scale of abscissae of this figure is, however, in terms of the path increment at one edge of the aperture taken with reference to the other edge, and our present Δ equals one-half of this. Thus, for example, the first minimum in the pattern is shown as occurring at 1.22λ, or when $\Delta = .61\lambda$.

The positions of the maxima and minima of the pattern may be ascer-

tained directly from the plotted curve, or we may locate these points by differentiating series (3′). This operation leads to a related series.*

Since the radius of the aperture, r, does not appear in the series (3′), it is obvious that the patterns for apertures of various radii are all similar. However, the radius, r, and the focal length, f, of the lens enter in determining the scale of abscissae of the graph when this scale is given in terms of z, the distance of a given field point from the center of the pattern. We have the relation:

$$\frac{z}{f}=\frac{\Delta}{r}\,. \tag{4}$$

Thus, for example, the first minimum will lie at the distance $z=f\Delta/r=.61\lambda f/r$ from the center of the pattern.

The series (3′) was first arrived at by G. B. Airy† in 1834. He tabulated its value for successive values of $2\pi\Delta/\lambda$ which he designated by the symbol n. He tabulated the series for values of $n=2\pi\Delta/\lambda=0$, .2, .4, etc., to $n=12$. Since Airy's time the problem has been attacked anew several times.‡ The series derived by Airy is closely related to the Bessel function known as J_1, which when n is used as the argument is written symbollicaly as $J_1(n)$. The part of Airy's series which is contained in the bracket of our equation (3′) is equal to $2J_1(n)/n$. Extensive tabulations of Bessel functions exist today which were not available in Airy's time.

We shall now show how the series (3′) is derived from the integral (3). We have:

$$S=2\int_0^r \sqrt{r^2-q^2}\cos\left(\frac{2\pi\Delta}{\lambda}\cdot\frac{q}{r}\right)dq=2r^2\int_0^r\sqrt{1-\frac{q^2}{r^2}}\cos\left(\frac{2\pi\Delta}{\lambda}\cdot\frac{q}{r}\right)\cdot d\frac{q}{r} \tag{5}$$

$$=2r^2\int_{w=0}^{w=1}\sqrt{1-w^2}\cos(nw)\cdot dw\,, \tag{5′}$$

where for abbreviation we have written w for q/r and n for $2\pi\Delta/\lambda$. We now introduce the well-known series expansion for the cosine term, obtaining:

$$S=2r^2\int_0^1\sqrt{1-w^2}\left[1-\frac{(nw)^2}{2}+\frac{(nw)^4}{\lfloor 4}-\frac{(nw)^6}{\lfloor 6}\cdots\right]dw\,. \tag{6}$$

* See, e.g., Gray, Mathews, and MacRobert, *A Treatise on Bessel Functions* (Macmillan, 1931), p. 191.

† *Trans. Camb. Phil. Soc.*, **5**, 283, 1834.

‡ Notably by E. Lommel, *Abhand. d. Bayer. Akad.*, **15**, 233, 1886. Lommel's treatment is in the main followed by Gray, Mathews, and MacRobert, *op. cit.*, chap. 14.

Each term may now be integrated separately. The fourth term, for example, becomes:

$$-2r^2 \int_0^1 \sqrt{1-w^2} \cdot \frac{(nw)^6}{\underline{6}}\, dw\,. \tag{7}$$

If we write $w = \sin x$ and then extend the integral from $x=0$ to $x=\pi/2$, this term assumes a form which permits of evaluation—and the same applies to the other terms. Introducing the aforementioned substitution, the term above becomes:

$$-\frac{2r^2n^6}{\underline{|6}} \int_{x=0}^{x=\pi/2} \cos^2 x \sin^6 x\, dx = -\frac{2r^2n^6}{\underline{|6}} \left[\int_0^{\pi/2} \sin^6 x\, dx - \int_0^{\pi/2} \sin^8 x\, dx \right]. \tag{8}$$

We now apply a well-established formula of integration concerning which more will be said presently. The formula in question is:

$$\int_0^{\pi/2} \sin^m x\, dx = \frac{1\cdot 3\cdot 5 \ldots . (m-1)}{2\cdot 4\cdot 6\cdot \ldots . m} \cdot \frac{\pi}{2}\,, \tag{9}$$

where m is an even integer. Applying this formula to the right-hand side of (8), we have:

$$-\frac{2r^2n^6}{\underline{|6}} \left[\frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6} \cdot \frac{\pi}{2} - \frac{1\cdot 3\cdot 5\cdot 7}{2\cdot 4\cdot 6\cdot 8} \cdot \frac{\pi}{2} \right] = -\frac{\pi r^2 n^6}{2^2\cdot 4^2\cdot 6^2\cdot 8}\,. \tag{10}$$

Remembering that $n = 2\pi\Delta/\lambda$, we see that the right-hand side of this equation is identical with the fourth term of the series (3′). The other terms of the series may be evaluated by a similar process.

The formula of integration (9) is given, for example, in B. O. Peirce's *Short Table of Integrals* (2d ed.), page 62, No. 483, and is not difficult to verify. It may be verified by applying formula No. 263, page 38, of the same table of integrals, namely:

$$\int \sin^m dx = -\frac{\sin^{m-1} x \cos x}{m} + \frac{m-1}{m} \int \sin^{m-2} x\, dx\,. \tag{11}$$

This formula may in turn be verified by differentiating the right-hand side, this process giving the clue to the manner of derivation of the formula.

APPENDIX D

THE ABERRATION OF THE CONCAVE GRATING

(Continuation of Chap. 6, Sec. 25)

We shall discuss the two most important special cases:

Case A. Field point at the center of curvature.—When the field point lies at C (Fig. 266), the optical paths from various points of the grating to the field point are all equal—there are no path differences and no path errors introduced upon leaving the grating. The path differences and path errors are limited to those which occur at incidence, and it is necessary to discuss these only. The condition in question is practically fulfilled in the Rowland

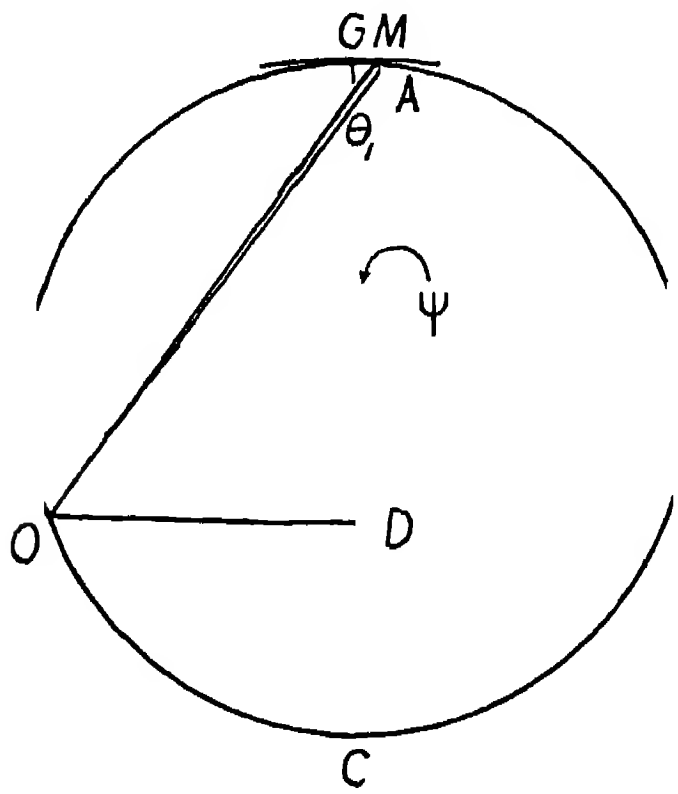


FIG. 266

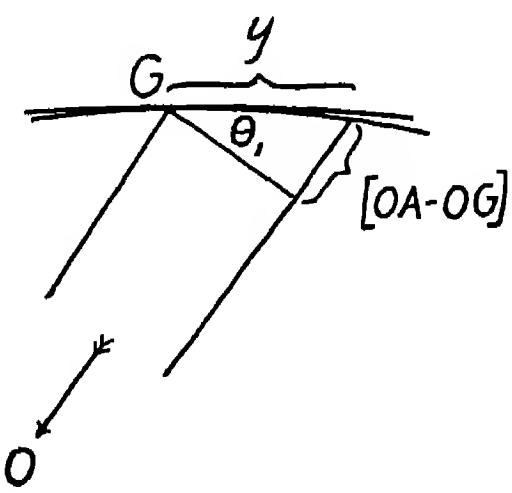


FIG. 267

mounting and in the Abney mounting.* Let us evaluate the aberration which exists when the grating is ruled in the usual manner with equal spaces along the chord. The notation of chapter 6, sections 23–25, will be used. The present Figure 266 is the same as Figure 131 of chapter 6. Equation (12) of chapter 6 may be written:

$$a' = \frac{a}{\cos\psi} = a\left(1 + \frac{AM}{CA}\right). \tag{1}$$

This equation holds accurately—there are no approximations involved.

The *theoretically correct* grating spacing, as measured along the arc, will be designated by the symbol a''. The value of a'' at a given point of the grating surface is determined by the fact that it must satisfy the equation

* Concave-grating mountings are described in chap. 6, secs. 28–33.

$a'' \sin \theta_1' = a \sin \theta_1$. Remembering that the angle OAC equals θ_1 and OMC equals θ_1', we now have:

$$(2) \qquad a'' = \frac{a \sin \theta_1}{\sin \theta_1'} = a \frac{OM}{OA} = a \left[1 + \frac{(OM - OA)}{OA} \right] .$$

It was shown in section 24 of chapter 6 that the small term $(OM - OA)/OA$ is approximately equal to AM/CA. In order to evaluate the aberration we must seek a higher approximation for this small term.

We have the identity $OA = OG + [OA - OG]$. Referring to Figure 267, the distance $[OA - OG] = y \sin \theta_1$ very nearly, therefore $OA = OG + y \sin \theta_1$. This value of OA is evidently a much closer approximation than the value $OA = CA \cos \theta_1$ previously used (see top of page 152); the previously used approximation amounted, practically, to saying $OA = OG$, because $OG = CG \cos \theta_1$ accurately, and $CG = CA$ very nearly. The new approximation is thus much closer.

Fortunately our previous approximation for the difference $(OM - OA)$, see Figure 132, page 151, namely $(OM - OA) = AM \cos \theta_1$, is still amply accurate and requires no correction.

In accordance with the foregoing:

$$(3) \qquad \frac{(OM - OA)}{OA} = \frac{AM \cos \theta_1}{OG + y \sin \theta_1} = \frac{AM \cos \theta_1}{R \cos \theta_1 + y \sin \theta_1} .$$

The term $R \cos \theta_1$ is much greater than either $AM \cos \theta_1$ or $y \sin \theta_1$. In this circumstance the right-hand side of this equation may be written approximately:

$$(3') \qquad \frac{AM \cos \theta_1}{R \cos \theta_1} - \frac{AM \cos \theta_1 \times y \sin \theta_1}{(R \cos \theta_1)^2} = \frac{AM}{R} - \frac{AM\, y}{R^2} \tan \theta_1$$

$$= \frac{AM}{CA} - \frac{AM\, y}{R^2} \tan \theta_1 .$$

The last is the desired evaluation of the term $(OM - OA)/OA$. Inserting it in equation (2), we have:

$$(4) \qquad a'' = a \left[1 + \frac{AM}{CA} - \frac{AM\, y}{R^2} \tan \theta_1 \right] .$$

Subtracting this equation from equation (1):

$$(5) \qquad a' - a'' = \frac{a\, AM\, y}{R^2} \tan \theta_1 .$$

The distance AM may be evaluated in terms of y by applying the sagittal relation in succession to the reference circle and to the grating surface. This relation yields $AM = y^2/2(R/2) - y^2/2R = y^2/2R$. Therefore:

$$a' - a'' = \frac{ay^3}{2R^3} \tan \theta_1 . \tag{6}$$

The presence of θ_1 in this equation shows that a given grating can be theoretically correct for only a given position of the slit. The error of the actual grating is small, however, except when θ_1 is large. For example, when the angle of incidence is less than 60°, the tangent has a value less than 2, and we have $a' - a'' < (ay^3/R^3) = a\psi^3$. Thus at the edge of the grating where $\psi = .01$, the difference between a' and a'' is less than one one-millionth of a grating space. We previously learned that the variation in a' in passing from the center of the grating to the edge is one-twenty-thousandth of a grating space. Thus, by ruling with equal spacing along the chord, the theoretically required variation of the grating space along the arc is taken care of with an error of less than $(1 \times 10^{-6})/(1/20{,}000) = .02$ of the total variation. When the grating space is constant along the chord, the variation along the arc is slightly greater than it should be for theoretically correct ruling. This is shown by the fact that $a' - a''$ is a positive quantity.

We shall now determine the difference between the position of a given line of the grating as determined by ruling with constant spacing along the chord and the position of the same line according to theoretically correct ruling. This "position error" is equal to the sum of the differences $(a' - a'')$ taken from the center of the grating to the point in question, or is, in other words, the integral of $(a' - a'')$ with reference to the running number of the line which we shall designate by ν. We thus have according to equation (6):

$$\int_0^{\nu} (a' - a'')\, d\nu = \frac{\tan \theta_1}{2R^3} \int_0^{\nu} y^3 a \, d\nu = \frac{\tan \theta_1}{2R^3} \int_0^{y} y^3 dy = \frac{y^4 \tan \theta_1}{8R^3} . \tag{7}$$

Employing this expression, let us evaluate the position error for a line near the edge of the grating, $y = 6.40$ cm, for a chosen angle of incidence, namely, $\theta_1 = 26°34'$, this being the angle for which the tangent has a value of 1/2. Then:

$$\frac{y^4 \tan \theta_1}{8R^3} = \frac{(6.40)^4}{16 \times (640)^3} = \frac{6.40 \times 10^{-6}}{16} = 4 \times 10^{-7} \text{ cm.}$$

Expressed in the fraction of a grating space of a grating having 20,000 lines per inch, this error is $(4 \times 10^{-7})/(1.27 \times 10^{-4}) = .003$. This shows to what de-

gree equal spacing along the chord approximates to the theoretically correct ruling so long as the angle of incidence is not all too large. The angle $\theta_1 = 26°34'$ corresponds to a wave-length of 5700 Å in the first order. In this order the allowable position error is .25 grating space. There is thus a factor of safety of about 80.

In passing to greater wave-lengths and higher spectral orders the initially large factor of safety becomes progressively smaller. The position error increases with increasing angle of incidence, and at the same time the permitted error decreases with increasing spectral order. The permitted error is $a/4m$, where m is the order. If we wish to have the third order of the middle violet, $\lambda = 4100$ Å, diffracted to our field point at the center of curvature, the required angle of incidence is about 75.5° and $\tan 75.5° = 3.87$. The position error at the edge of the grating is then $[(6.40)^4 \times 3.87]/[8(640)^3] = 3 \times 10^{-6}$ cm, or $(3 \times 10^{-6})/(1.27 \times 10^{-4}) = .024$ grating space. The permitted error in the third order is one-twelfth of a grating space—the factor of safety is now less than 4.

The permitted width of a grating under given conditions may be readily computed. It decreases with increasing wave-length and spectral order. The position error is, by equation (7), equal to $y^4 \tan \theta_1/8R^3$ and the permitted error at the edge of the grating is $a/4m$. Therefore, if y is the permitted half-width of the grating:

$$\frac{y^4 \tan \theta_1}{8R^3} = \frac{a}{4m},$$

or

$$y = \sqrt[4]{\frac{2aR^3}{m \tan \theta_1}}. \tag{8}$$

From this equation the following permitted widths may be deduced for a 21-foot concave grating of 20,000 lines per inch.* For $\lambda = 5000$ Å in the first order θ_1 is 23° and the permitted width is 40 cm. By way of comparison the actual width is usually between 12 and 14 cm. For $\lambda = 5000$ Å in the second order θ_1 is 52° and the permitted width is 25.5 cm. When observations are made near the center of curvature, as we are here supposing, the third order of 5000 Å cannot be reached with a 20,000-line-per-inch grating, but for the third order of the above mentioned wave-length of 4100 Å the angle $\theta_1 = 75.5°$ and the permitted width is 17.5 cm.

Case B. Field point coinciding with slit.—In the Eagle mounting the photographic plate is situated just above or just below the slit. From our

* $R = 640$ cm; $a = 1.27 \times 10^{-4}$ cm.

present point of view we may say that the field point coincides with the slit and that the angle of diffraction is equal to the angle of incidence. In this case the grating law becomes $m\lambda = 2a \sin \theta_1$. The errors in optical path which occur at incidence are the same for a given angle θ_1 as those which occur under Case A, when the field point lies at the center of curvature, but now, equal path errors are added upon leaving the grating. However, both path errors would be corrected simultaneously by a given shift in the positions of the lines of the grating, consequently the position errors computed under Case A hold identically in Case B for a given angle θ_1. The difference is that for a given angle either the observed wave-length is doubled or the spectral order is doubled in Case B. For a given wave-length and spectral order the angle is less than half as great in Case B and thus the position error is less than half as great. The permitted width is the same in Case B as in Case A when the angle and spectral order are the same. In other words, the permitted width is the same for double the wave-length. The Eagle mounting thus has a distinct advantage in regard to aberration because it involves the use of smaller angles for a given wave-length and spectral order.

For work in the extreme ultra-violet, angles of incidence and diffraction approaching the grazing angle are now frequently used. The spectrum is observed, not as in the Eagle mounting on the same side of the normal as the incident beam, but on the opposite side of the normal. When, as in such work, nearly grazing angles are involved, the optimum width of the grating may be much less than the actual width. This case has recently been discussed by Mack, Stehn, and Edlén,* and by I. S. Bowen.† The optimum width of the grating may be but little more than 1 cm.

* *JOSA*, **22**, 245, 1932. † *Ibid.*, **23**, 313, 1933.

APPENDIX E

THE ASTIGMATISM OF THE CONCAVE GRATING

(Continuation of Chap. 6, Sec. 27)

Let us suppose that a ray of light OM (Fig. 268) is incident upon *a single groove*, GH, in a reflecting surface. This surface, through GH, may or may not be perpendicular to the plane of our figure. Erect a plane MC perpendicular to the groove at the point of incidence, M, and denote the angle which the ray, indicated as being incident from below, makes with this plane, by α. Light will be diffracted or scattered in all directions which make an angle α above the plane MC. The direction of diffraction generates a circular cone as illustrated. The elements of the cone make an angle α with the plane MC. That this is true may be shown by applying Huygens' construction, every point along the groove being considered as the center of a Huygens wavelet of appropriate radius.

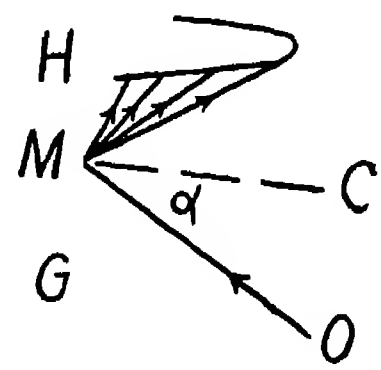


FIG. 268

When a beam is incident upon a *grating*, the diffracted beam of order n and wave-length λ leaves each point of the grating along a particular el ment of the aforementioned cone erected at that point—along the eleme along which the grating law is satisfied. The diffracted beam thus makes angle α with the plane which is perpendicular to the grooves at the poin incidence.

We shall now consider a vertical fan of rays originating from O (Fig 9) incident upon the grating G. The rays diffracted by the grating co e al another vertical fan. This fan i lies in a different vertical plan gh G, but if we may be permitte ver- look this fact for just a mome may suppose that a given spectru is located at the distance GP fro rating. If the point M be considered at the upper end of the ruled lines of grating, the diffracted fan will have a height P_0P'' which will be the th of the vertical astigmatic focal line. Alternatively, we may say tha PP'' represents the half-length of the focal line. Let us now evaluate th half-length, taking due account of the fact that the incident vertical fa and the diffracted one in general lie in different vertical planes throu G.

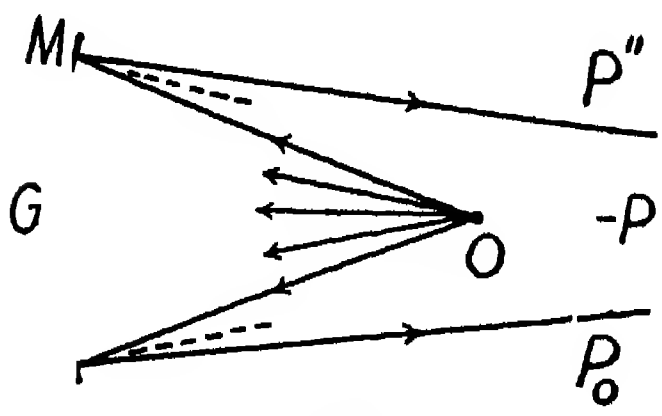


FIG. 269

Referring to Figure 270, the upper portion of this figure represents a section in the XY-plane, a horizontal section in the plane of the Rowland circle.* The lower portion represents a section in the XZ-plane, that is, a vertical section through the center of the grating, G, and the center of curvature, C. The point source, O, and the field point, P, lie on the Rowland circle and these points project into the XZ-plane at the positions indicated. Suppose that M is a point at a height z above the center of the grating

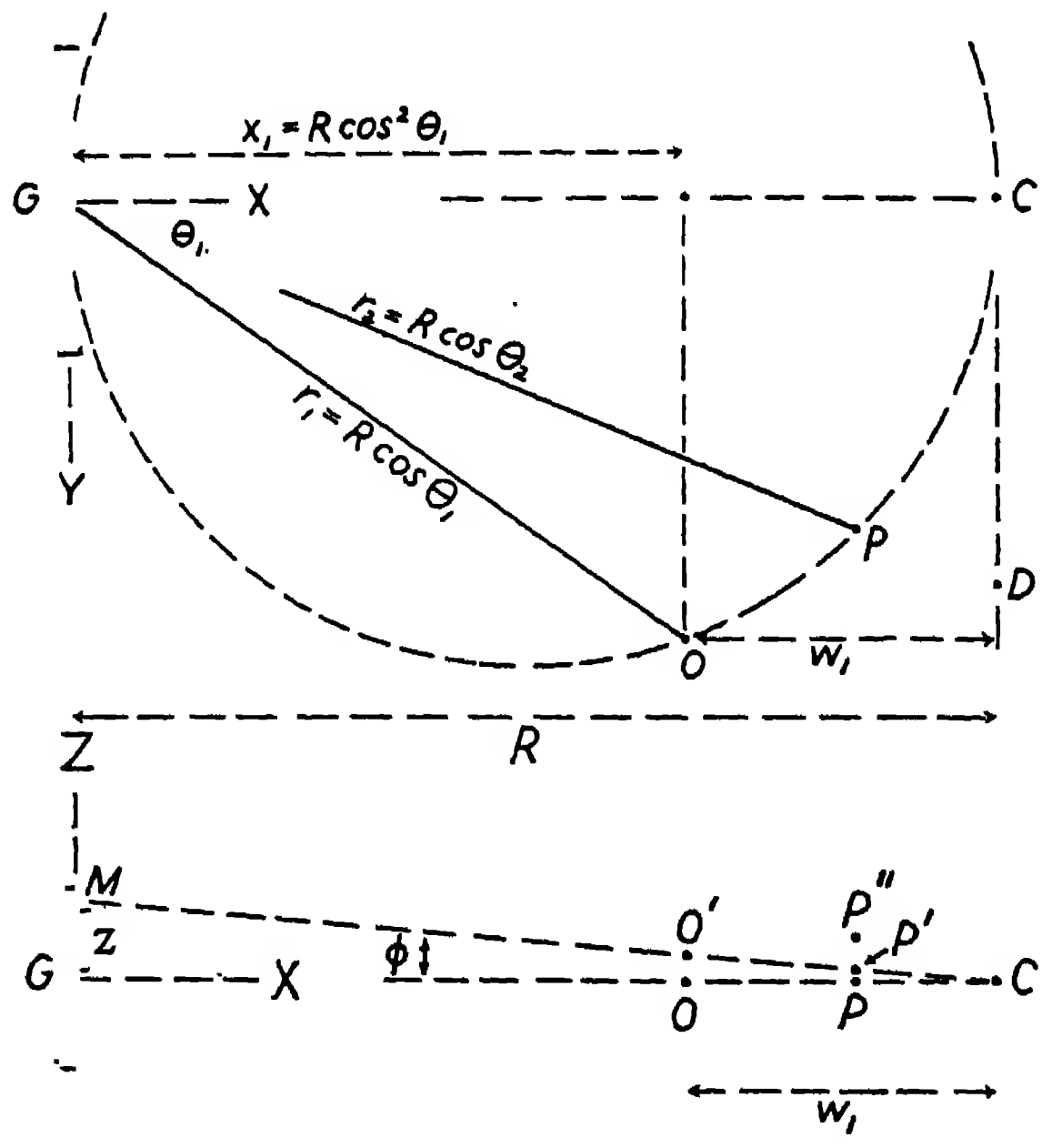


FIG. 270

surface. The plane which is perpendicular to the surface at M, and also perpendicular to the ruled line, is MC. The ray from O which is incident at M makes an angle α with this plane (not indicated in this figure). If this ray were projected into the XZ-plane, the angle which the projection makes with the plane MC would differ from α, the angle would be greater, and for this reason the ray and the angle are not indicated in the lower part of the figure. The diffracted ray leaves the grating above the plane MC at the angle α with this plane. This angle would also be altered by projection and is therefore not indicated. The diffracted ray passes above the point P through a point indicated by P''. The length PP'' projects into the vertical plane without alteration. Moreover, if M is considered to lie at the upper end of the ruled lines of the grating, the length PP'' will be the half-length

* Axes of coordinates being taken as stated in chap. 6, sec. 27.

of the vertical astigmatic focal line, that is, the half-length of the spectrum line, which we shall denote by L. We now have:

$$PP'' \equiv L = PP' + P'P''. \tag{1}$$

The two parts on the right will be separately evaluated.

The distance PP' is the distance from the field point on the Rowland circle to the plane MC. Similarly, the distance OO' is the distance from the source point to this plane. These two distances play similar rôles in our theory. Let us obtain an evaluation for OO', which we shall also need, and we shall then be able to write an expression for PP' by analogy. The subscript 1 will refer to the source and the subscript 2 will refer to the field point. The distance r_1 equals $R \cos \theta_1$, and the X-coordinate of the source is $x_1 = r_1 \cos \theta_1 = R \cos^2 \theta_1$. Referring to the figure for the explanation of the meanings of other symbols and introducing evident approximations for the vertical angles, which are small, we have:

$$\frac{z}{R} = \phi = \frac{OO'}{w_1} = \frac{OO'}{R - x_1} = \frac{OO'}{R - R \cos^2 \theta_1} = \frac{OO'}{R \sin^2 \theta_1}. \tag{2}$$

The term at the left and the term at the right yield:

$$OO' = z \sin^2 \theta_1 \tag{3}$$

and by analogy:

$$PP' = z \sin^2 \theta_2 . \tag{4}$$

Furthermore, $OO'/r_1 = a = P'P''/r_2$, whence, remembering that $r_1 = R \cos \theta_1$ and $r_2 = R \cos \theta_2$, and employing equation (3):

$$P'P'' = \frac{r_2}{r_1} OO' = \frac{\cos \theta_2}{\cos \theta_1} (z \sin^2 \theta_1) . \tag{5}$$

By equations (1), (4), and (5) the half-length of the spectrum line

$$L = z \left[\sin^2 \theta_2 + \frac{\cos \theta_2}{\cos \theta_1} \sin^2 \theta_1 \right] . \tag{6}$$

This is our desired result, applying when a point source of light is used. It is to be noted that the length of the spectrum line is independent of the radius of curvature of the grating.*

* Curves giving the length of the spectrum line as a function of, e.g., the angle of incidence, have recently been published by G. H. Dieke, *JOSA*, **23**, 274, 1933.

Let us now suppose that instead of using a point source of light we use a vertical slit which extends a distance s above the Rowland circle and an equal distance s below this circle, that is, s is the half-length of the slit. A point source at the bottom of the slit will give rise to a vertical-line image of which the center will lie at a height sr_2/r_1 above the Rowland circle. Accordingly, when a slit is used, the half-length of the spectrum line is increased by the amount $sr_2/r_1 = s \cos \theta_2/\cos \theta_1$. Denoting the half-length of the spectrum line when a slit is used by L_s, we have:

$$L_s = z\left[\sin^2 \theta_2 + \frac{\sin^2 \theta_1 \cos \theta_2}{\cos \theta_1}\right] + \frac{s \cos \theta_2}{\cos \theta_1}. \tag{7}$$

When z represents the half-length of the ruled lines of the grating, then L in equation (6) represents the half-length of the spectrum line. Alternatively, when z represents the full length of the ruled lines, then L represents the full length of the spectrum line. Similar conclusions apply to equation (7), where, however, in addition s must in the latter case be interpreted as representing the full length of the slit.

We may consider L_s in equation (7) to be made up of two parts, the first part represented by the first term on the right, arising from astigmatism, and the second part represented by the second term, arising from slit length. If all of the intensity originally lost due to astigmatism is to be regained, then the second term must be at least equal to the first term; see again chapter 6, section 27, especially the discussion apropos of Figure 136. With this principle as a basis, the length of slit which is necessary to regain all of the lost intensity may be readily calculated. The use of a still longer slit would add nothing further to the intensity. When, on the other hand, it is not feasible to use a slit which is sufficiently long to regain all of the lost intensity, then, introducing the value of the maximum slit length which it is feasible to use, we may, according to the same principles, calculate how great a length of the ruled lines is effective. The portions of the ruled lines which lie above and below this central effective portion contribute nothing further to the intensity.

We shall now consider the astigmatism proper, as given by equation (6), in two important special cases:

1. When the field point lies at the center of curvature.—The angle of diffraction θ_2 now equals zero and the grating law becomes $m\lambda = a \sin \theta_1$ and thus $\sin \theta_1 = m\lambda/a$ and $\cos \theta_1 = \sqrt{1-(m^2\lambda^2/a^2)}$. Equation (6) thus becomes:

$$L' = \frac{zm^2\lambda^2}{a^2\sqrt{1-\dfrac{m^2\lambda^2}{a^2}}}. \tag{8}$$

From this equation we may calculate the astigmatism which obtains in a spectrum photographed in the vicinity of the center of curvature, as in the Rowland mounting or the Abney mounting (see sections on concave-grating mountings, chap. 6, secs. 28 ff.). The foregoing equation is in a form which is convenient for comparing the astigmatism which obtains in a given wave-length and spectral order, when the Rowland or Abney mounting is used, with that which obtains when some other form of mounting is used. Such a comparison will be made presently.

2. *When the field point coincides with the source.*—We now have $\theta_1 = \theta_2$ and the grating law becomes $m\lambda = 2\ a \sin \theta_1$. Substituting in equation (6) for $\sin \theta_1$ and $\sin \theta_2$ the value $m\lambda/2a$, we have:

$$L'' = \frac{zm^2\lambda^2}{2a^2}. \tag{9}$$

When the Eagle mounting is used the spectrum is photographed in the vicinity of the source and we may use the foregoing equation to calculate the astigmatism.

The right-hand side of equation (8) is always more than twice as great as the right-hand side of equation (9), consequently the astigmatism is more than twice as great, for a given wave-length and spectral order, when the Rowland or Abney mounting is used than when the Eagle mounting is used. Consequently, the Eagle mounting has a decided advantage in regard to intensity when it is not feasible to use a slit of sufficient length to regain all of the intensity lost due to astigmatism, as is so often the case.

In conclusion we shall evaluate the distance from the center of the grating to the point where a vertical fan of rays diffracted by a central vertical strip of the grating is brought to a focus. This distance will be denoted by r_2'. We shall consider two cases:

Case A. When the source (a point source) occupies *a general position on the Rowland circle:* Referring to the upper portion of Figure 270, let us suppose that the line GP, extended, intersects the line CD at D; then $GD = R/\cos \theta_2$. Now take a vertical section through G, P, and D, as represented in Figure 271. The plane of this figure intersects the plane erected at M perpendicular to the ruled lines, along the line MD, and D is the [illegible] where this intersection meets the XY-plane. Denote the angle wh[illegible]

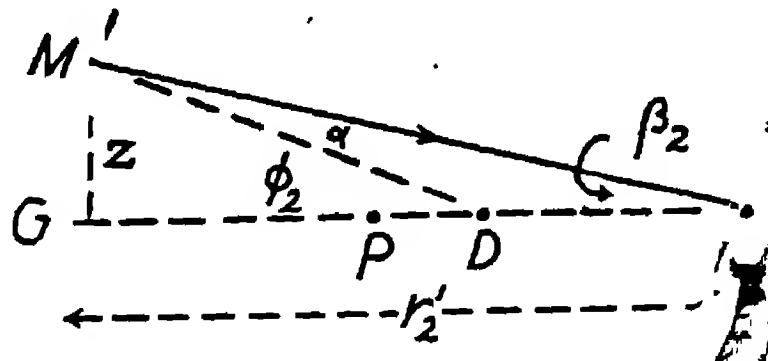


FIG. 271

makes with the XY-plane by ϕ_2, and the angle which the diffracted ray makes with the XY-plane by β_2; then:

$$\frac{z}{r_2'}=\beta_2=\phi_2-\alpha \qquad \text{or} \qquad r_2'=\frac{z}{\phi_2-\alpha}.$$

Now:

$$\phi_2=\frac{z}{GD}=\frac{z\cos\theta_2}{R},$$

and invoking Figure 270 and equation (3):

$$\alpha=\frac{OO'}{r_1}=\frac{z\sin^2\theta_1}{R\cos\theta_1}.$$

Therefore, finally:

$$(10) \qquad r_2'=\frac{z}{\phi_2-\alpha}=\frac{z}{\dfrac{z\cos\theta_2}{R}-\dfrac{z\sin^2\theta_1}{R\cos\theta_1}}=\frac{R}{\cos\theta_1\cos\theta_2-\sin^2\theta_1}.$$

This is the desired expression for the distance from the grating to the point at which the vertical diffracted fan comes to a focus, or, what amounts to the same, it is the distance from the grating to the *horizontal* focal line which is formed in a given wave-length and spectral order.

Case B. When the source (again a point source) lies in the XY-plane *at infinity*, in a general direction: The incident rays are parallel and horizontal. This case is realized in practice by placing the source at a finite distance and collimating the incident beam by means of a lens or mirror. Designate the angle of incidence by θ_1. We shall consider only the special case in which the angle of diffraction, θ_2, equals zero, that is, when the field point lies on the normal to the grating—on the X-axis. Referring to Figure 272, which is drawn in the XZ-plane, the field point in question is Q, and C is the center of curvature of the grating. The beam is incident horizontally at an angle θ_1 with the plane of our figure. The ray which is incident at M makes an angle α with the plane MC. The ray diffracted at M leaves at this same angle α below the plane MC and meets the X-axis at Q, which lies at the distance r_2' from the grating. Denote the angle made by the diffracted ray with the X-axis by β_2. Now $z/r_2'=\beta_2=\alpha+\phi$, or:

$$(11) \qquad r_2'=\frac{z}{\alpha+\phi}=\frac{z}{\alpha+\dfrac{z}{R}}.$$

The angle α remains to be evaluated. Consider a vertical plane taken through the center of the grating in the direction of the source. This plane intersects our plane MC and this intersection makes an angle with the XY-plane which we shall call ϕ_1. Since the incident ray lies above the XY-plane and is parallel to it, this ray makes the same angle ϕ_1 with the plane MC.

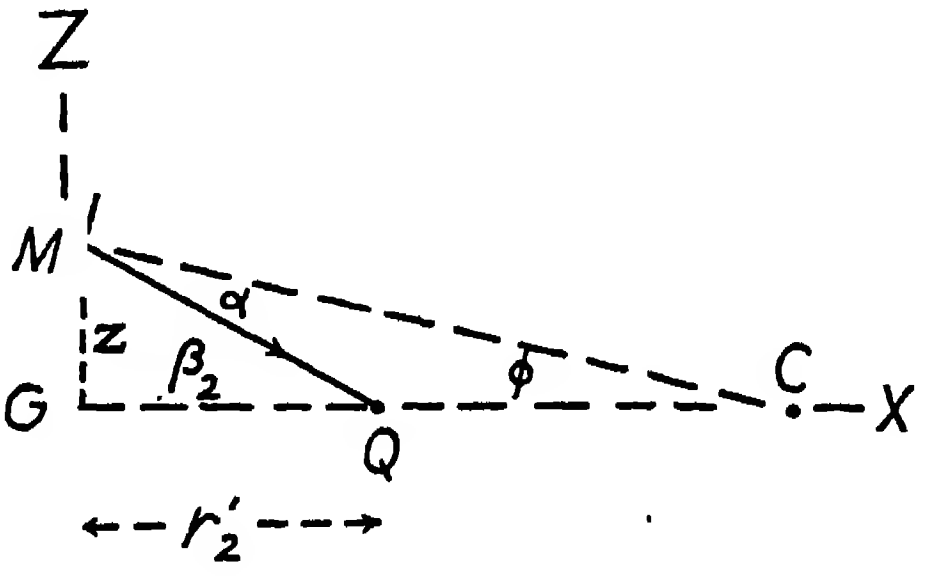


FIG. 272

Therefore, $\alpha=\phi_1$ and, moreover, $\phi_1=\phi \cos \theta_1=(z/R) \cos \theta_1$. Hence $\alpha=(z/R)\cdot\cos \theta_1$ and substituting this value in equation (11):

$$r_2'=\frac{z}{\frac{z}{R}\cos \theta_1+\frac{z}{R}}=\frac{R}{\cos \theta_1+1}. \tag{12}$$

This is the desired evaluation of r_2' when the incident beam is parallel and the diffracted vertical fan is the one which is diffracted at an angle $\theta_2=0$, or along the normal to the grating. Now it will be shown presently that this value of r_2', or the distance at which a diffracted vertical fan comes to a focus, is equal to the distance r_2 at which a diffracted horizontal fan comes to a focus under the same condition of incidence and for the same angle of diffraction $\theta_2=0$. The field point at which a horizontal fan comes to a focus we have previously designated by P, and this is the point at which the spectral image is observed or photographed. Thus, in the present case the points P and Q coincide, or, in other words, a point source of light gives rise to a point image—there is no astigmatism—the concave grating in this case yields stigmatic spectral images. Practically speaking, the spectrum is stigmatic over an appreciable range on either side of the normal and it is within this range that the spectrum is photographed when the so-called stigmatic mounting of the concave grating is used (see chap. 6, sec. 32).

It remains to prove the fact stated above that in the case under consideration $r_2=r_2'$. This follows immediately from the general focal relation developed in chapter 6, section 26, in particular from equation (19). Divide

both the numerator and the denominator of this equation by r_1 and then place $r_1=\infty$ and $\cos\theta_2=1$. Then $r_2=R/(\cos\theta_1+1)=r_2'$.

The treatments of the astigmatism of the concave grating which have come to the author's notice are as follows: C. Runge, in Kayser's *Handb. der Spekt.*, **1**, 460; Runge and Mannkopf, *Zeit. f. Phys.*, **45**, 13, 1927; J. L. Sirks, *Astron. and Astrophys.*, **13**, 863, 1894; A. Eagle, *Astrophys. Jour.*, **31**, 121, 1910, in particular pp. 130 ff.; and S. A. Mitchell, *Johns Hopkins Univ. Circulars*, **17**, 56, 1898; the last is marred by typographical errors and other faults.

Since writing the foregoing, an article by G. H. Dieke has appeared (see *JOSA*, **23**, 274, 1933).

APPENDIX F

ROWLAND GHOSTS

(Continuation of Chap. 6, Sec. 37)

We shall begin by showing that when the periodic error is a pure simple harmonic one having the period of the screw, the ghosts of the third order will vanish. Our previous Figure 151 is here repeated as Figure 273. The vector A' is for a groove of the actual grating, the vector A is for the same groove of the ideal grating, and v is the effect of the error of position of the groove. Small vectors such as v are to be erected perpendicularly along the

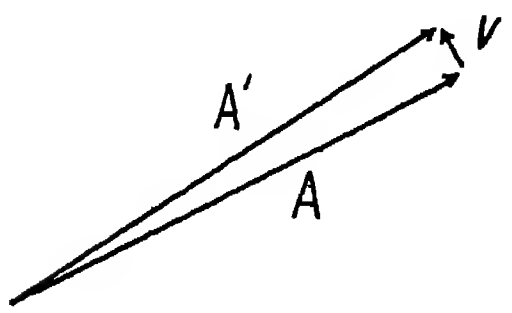


FIG. 273

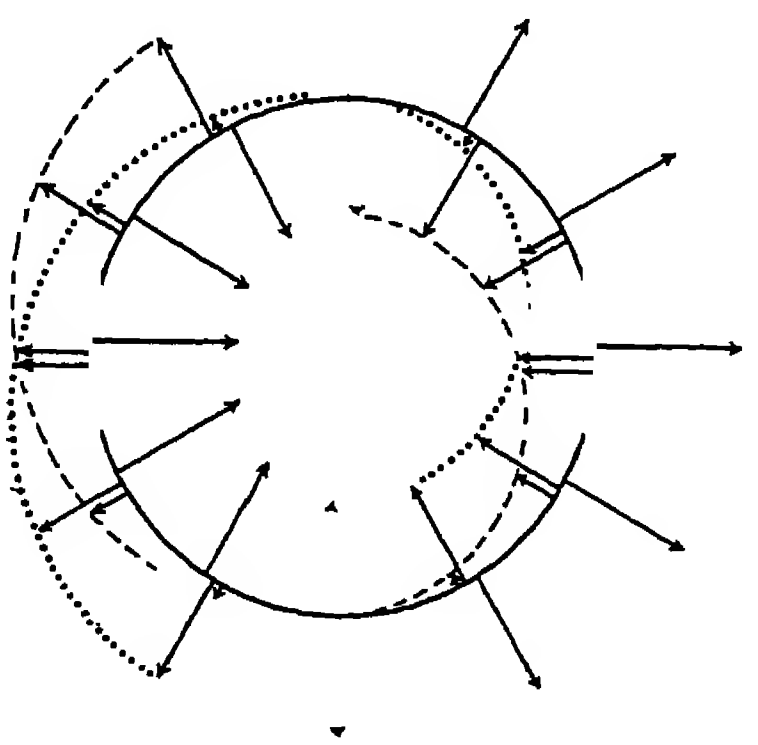


FIG. 274

vibration polygon. When the polygon makes three convolutions per turn of the screw, the small vectors v, much exaggerated in length for purposes of representation, will appear as shown in Figure 274. Our problem is to show that the sum of these perpendicular vectors is zero. Let us consider in turn the sum of the vertical components and the sum of the horizontal components.

That the sum of the vertical components is zero follows from symmetry considerations, as is evident from Figure 274. For every vector pointing upward at a certain angle there is a vector of equal length pointing downward at the same angle. The fact can also be proved by steps analogous to those which we shall carry through in detail for the horizontal components.

In regard to the horizontal components: Referring to Figure 275, the horizontal component of any small vector v is:

$$v_h = v \sin \psi , \tag{1}$$

where ψ is the angle which the tangent to the vibration curve for the ideal grating makes with the horizontal. If the error is simple harmonic, the value of v will be given by the equation:

$$v = v_0 \sin \phi , \tag{2}$$

where v_0, a constant, is the maximum length which any of the small vectors v may have and this is directly proportional to the amplitude of the error of position of the grooves, and ϕ is the angle through which the spacing head of the ruling machine had rotated when ruling the line in question, this angle being measured from the position where the lines of the actual grating are correctly located. If there are N lines ruled per turn of the spacing head, we have $\phi = 2\pi(\nu/N)$, where ν is the running number of the line in question, counted from the same zero from which ϕ is measured. Now, when we have three convolutions of the vibration polygon per turn of the screw, the value of ψ for the groove of number ν is $3(2\pi\nu/N)$. In other words, $\psi = 3\phi$. The sum of the horizontal components of the small vectors is therefore according to equations (1) and (2):

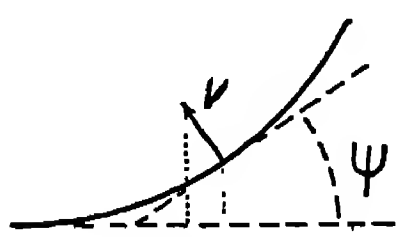


FIG. 275

$$\sum v_h = \sum_{\psi=0}^{\psi=6\pi} v \sin \psi = v_0 \sum_{\phi=0}^{\phi=2\pi} \sin \phi \sin 3\phi . \tag{3}$$

The number of lines ruled per turn of the screw is large, and under this circumstance the foregoing summation will be equal to zero. This follows from the fact that the corresponding integral, namely, $\int_0^{2\pi} \sin \phi \sin 3\phi \, d\phi$, is equal to zero. This integral is a special case of the integral $\int_0^{2\pi} \sin j\phi \sin k\phi \, d\phi$, where j and k are integers. This integral is zero whenever j differs from k, as it does in the present case.* Accordingly, we may conclude that

* We have $\sin j\phi \sin k\phi = -\frac{1}{2} \cos (j+k)\phi + \frac{1}{2} \cos (j-k)\phi$. The first term integrated from 0 to 2π is zero at both limits in any case, i.e., both when j equals k and when it differs from k. The second term is zero at both limits only when j differs from k.

When j equals k, $\cos (j-k)\phi = 1$, and hence $\int_0^{2\pi} \frac{1}{2} \cos (j-k)\phi d\phi = \pi$.

the sum of the horizontal components of the small vectors is zero, and hence the third-order ghosts vanish.

It may be shown in the same way that the ghosts of all orders higher than the first will vanish when the error is simple harmonic.

We shall now derive an equation for the intensity of the first-order ghosts: If the periodic error is simple harmonic with the period of the screw—that is, with a period of N lines—the error of position e, measured as a fraction of a centimeter or inch, of the line of running number ν will be:

$$e = b \sin \phi = b \sin \frac{2\pi\nu}{N}, \tag{4}$$

where b is a constant—the amplitude of the error. The disturbance contributed by the line ν arrives, in the spectrum of order m, with an error of phase angle of $2\pi m(e/a)$, where a is the grating space. The angle in question is the one between the vectors A and A' (Fig. 273). When this angle is small we may write:

$$2\pi m \frac{e}{a} = \frac{v}{A}, \text{ or } v = \frac{2\pi m A}{a} e. \tag{5}$$

The intensity of the first-order ghosts, in the spectrum of order m, is $(n'ms_1)^2$, where as before n' is the number of turns of the screw involved in ruling the grating and s_1 is the error of closure for one convolution of the vibration polygon when it makes one convolution for each turn of the screw, in the first-order spectrum. We must evaluate s_1 in suitable terms. In the first-order spectrum, ψ is equal to ϕ, hence:

$$s_1 = \sum v_h = \sum_{\psi=0}^{\psi=2\pi} v \sin \psi = \sum_{\phi=0}^{\phi=2\pi} v \sin \phi. \tag{6}$$

We now substitute the value of v from equation (5), remembering that $m=1$, since s_1 is for the first-order spectrum. We then in turn substitute the value of e from equation (4). Thus:

$$s_1 = \frac{2\pi A}{a} \sum_0^{2\pi} e \sin \phi = \frac{2\pi A b}{a} \sum_0^{2\pi} \sin^2 \phi. \tag{7}$$

The increment of angle of rotation of the screw, between ruling successive lines, is $\Delta\phi = 2\pi/N$, or $(N/2\pi)\Delta\phi = 1$. We may thus multiply the right-hand side of the foregoing equation, outside of the summation sign by $N/2\pi$ and inside of it by $\Delta\phi$ without destroying the equality, and, since the number of

lines ruled per turn of the screw is large, we may express the summation as an integral. Hence:

$$s_1 = \frac{NAb}{a} \sum_0^{2\pi} \sin^2 \phi \, \Delta\phi = \frac{NAb}{a} \int_0^{2\pi} \sin^2 \phi \, d\phi \, . \tag{8}$$

The value of the integral is π, as shown in the previous footnote. Hence:

$$s_1 = \frac{\pi NAb}{a} \, . \tag{9}$$

Therefore the intensity of the ghosts in the spectrum of order m becomes:

$$I_g = (n'ms_1)^2 = \left(\frac{n'm \, \pi \, NAb}{a} \right)^2 . \tag{10}$$

Now $(n'NA)^2 = (nA)^2 = I_p$, the intensity which the parent-line would have if the grating were an ideal one. Actually the intensity of the parent-line will differ only inappreciably from this. Therefore equation (10) becomes:

$$I_g = I_p \left(\frac{\pi mb}{a} \right)^2 . \tag{11}$$

This is the desired expression for the intensity of the first-order ghosts The intensity increases with the square of the amplitude of the error.

If the error is not simple harmonic we may express it as a Fourier series; that is, we may write for the error of position of any ruled line:

$$e = b \sin \phi + c \sin 2\phi + d \sin 3\phi + \cdots , \tag{12}$$

where as before $\phi = 2\pi\nu/N$. If we introduce this value of e into equation (7) and then proceed in the same manner as previously, we will again arrive at equation (11) for the intensity of the first-order ghosts, because the new terms for e, as given by equation (12), give rise to definite integrals of which the value taken between the limits 0 and 2π is equal to zero. Hence the new terms drop out.

The intensities of the second- and third-order ghosts may be shown to be spectively:

$$\left(\frac{\pi mc}{a} \right)^2 , \tag{13}$$

$$I_{g,3} = I_p \left(\frac{\pi md}{a} \right)^2 .$$

That is, the intensity of the ghost of each order is determined by the amplitude of the corresponding term of the Fourier series, equation (12). Thus the intensities of the ghosts provide an analysis of the nature of the periodic error.

We shall now show that the periodic displacement of any line from its correct position should not exceed a maximum value of about $a/100$: The intensity of the ghosts should not be more than 2 or 3 per cent, in the highest spectral order which is to be used. For the sake of definiteness, suppose the ghosts may have an intensity of 2.5 per cent in the fifth order, which means that in the first order they may have an intensity of only one-twenty-fifth of this, or one-one-thousandth. Applying equation (11), which is for the first-order ghosts, we have:

$$I_0 = \frac{I_p}{1000} = I_p \left(\frac{\pi b}{a}\right)^2 .$$

Taking π^2 equal to 10, this yields $b = a/100$. In the same way equations (13) lead to $c = a/100$ and $d = a/100$. If we allowed an error twice as great as this, the ghosts would be four times as intense and would be troublesome, especially in the higher spectral orders. The fundamental component of the error usually dominates, and hence $a/100$ will be, practically, about the upper limit for the permissible maximum error of position.

The value $a/100$ allows for a grating of 20,000 lines per inch, only one-two-millionth inch as the maximum error of position. Moreover, the maximum error is an accumulated error. It is the error which accumulates during one-quarter of a turn of the screw, in ruling 250 lines—and it must not exceed one-two-millionth inch!

The amount Δa, by which any one grating space differs from the correct spacing a, is much smaller still. This is obtained by differentiating the error of position e with reference to the line number, ν. Supposing that the error is simple harmonic, we have by equation (4):

$$\Delta a = \frac{de}{d\nu} = b \cos \phi \frac{d\phi}{d\nu} = \frac{2\pi b}{N} \cos \phi$$

and for the maximum permitted value of Δa:

$$(\Delta a)_{\max} = \frac{2\pi b}{N} .$$

When $N=1000$ and $a=1/20{,}000$ inch, this yields

$$(\Delta a)_{\max}=\frac{2\pi\,\dfrac{a}{100}}{1000}=\frac{a}{16{,}000}=\frac{1}{320{,}000{,}000}\text{ inch.}$$

That is, the grating space must not be permitted to vary *periodically*, with the period of the screw, by much more than this amount, and even this amount of variation can be readily detected by the resulting ghosts. It is thus not surprising that the effects of periodic error cannot be gotten rid of.

Corresponding to the terms $c \sin 2\phi$, $d \sin 3\phi$, etc., of equation (12), the values of $(\Delta a)_{\max}$, are $2a/16{,}000$, $3a/16{,}000$, etc. That is, the permitted values are greater. But since b is as a rule decidedly greater than c, d, etc., these larger permitted values for the harmonics are less likely to be reached than is the value $(\Delta a)_{\max}=a/16{,}000$ for the fundamental, and thus $a/16{,}000$ is, practically, the upper limit for the permitted periodic variation of the grating space, for a grating of 20,000 lines per inch.

APPENDIX G

THE DIFFRACTION OF X-RAYS

(Continuation of Chap. 8)

Intercept ratios and Miller indices.—Every complete set of net planes of a primary lattice includes all of the points of the lattice—no point is *omitted;* this follows because all points of the lattice are alike and have neighbors in the same relative positions, and all points must play similar rôles in the lattice.

While intercept ratios are sometimes designated by fractions, for example, $\frac{2}{3}:1:-\frac{3}{5}$, the middle number being then always taken as unity, it is more common practice to express them by integers, thus $10:15:-9$—and it is always possible to express them by integers. A given plane of a parallel set does not necessarily intersect each of the co-ordinate axes at an integral number of times the lattice constant, but some plane of the set of necessity does so. In fact, for one plane of the set, the intercepts, each expressed in terms of the corresponding lattice constant, will be numbers which are prime to each other; another plane will make double these intercepts; a third plane will make three times these intercepts, etc. The planes in question are not usually consecutive planes of the set, however.

Since the intercept ratios of any crystal plane are always expressible by integers and the Miller indices are related to the intercept ratios by the reciprocal relation explained in chapter 8, section 4, the Miller indices will also be expressible by integers. For planes having physical importance the indices are small integers, that is, they are "simple." The same is true of the numbers giving the intercept ratios; however, an infinite intercept, denoting parallelism to a corresponding co-ordinate axis, must also be regarded as simple. The net planes of all sets for which the indices involve large integers lie close together, as will appear later. These planes are sparsely occupied by lattice points and are unimportant physically, for reasons which were explained in chapter 8, section 5. The fact that the Miller indices of planes having physical importance are expressible by small integers is called "the law of rational indices." This law was discovered as the result of a multitude of measurements upon crystal faces, made many years before the discovery of the diffraction of X-rays by crystals, and the conception of the space lattice arose also many years before the discovery of X-ray diffraction as the simplest means of interpreting the law of rational indices.

If p, q, and r are the integers which represent the intercept ratios of a set

of net planes of a cubic crystal having a lattice constant a, one of these planes makes intercepts upon the co-ordinate axes equal to pa, qa, and ra; the equation of this plane is:

$$\frac{x}{pa}+\frac{y}{qa}+\frac{z}{ra}=1\,. \tag{1}$$

By putting the left side of this equation equal to zero, we obtain the equation of a parallel plane through the origin. In the resulting equation the coefficients of x, y, and z may be multiplied or divided by any common factor. Let us multiply by $pqra$. Thus, for the parallel plane through the origin:

$$qr\cdot x+rp\cdot y+pq\cdot z=0\,. \tag{2}$$

The Miller indices are defined by the equations:

$$h=\frac{qr}{m}\,,\qquad k=\frac{rp}{m}\,,\qquad l=\frac{pq}{m}\,, \tag{3}$$

where m is the greatest common divisor of the products qr, rp, and pq. Hence the equation of a plane through the origin in terms of its Miller indices is:

$$hx+ky+lz=0\,. \tag{4}$$

The direction cosines of a normal to this plane are:

$$\frac{h}{\sqrt{h^2+k^2+l^2}}\,,\qquad \frac{k}{\sqrt{h^2+k^2+l^2}}\,,\qquad \frac{l}{\sqrt{h^2+k^2+l^2}}\,, \tag{5}$$

and a normal erected at the origin passes through the lattice point having co-ordinates (ha, ka, la). This is the first lattice point through which the normal passes. It passes also through the lattice points having co-ordinates ($2ha$, $2ka$, $2la$), ($3ha$, $3ka$, $3la$), etc.

Distance between planes, in terms of indices.—The minimal or perpendicular distance between the successive net planes of a given set may be expressed as a function of the lattice constant a, and of the Miller indices (h k l) of the set of planes in question. This distance is the one for which the symbol D was used in giving Bragg's law, equation (3) of chapter 8. This symbol will be retained to denote the distance in question. The normal to the planes, erected at the origin, passes through the lattice point having co-ordinates (ha, ka, la). Figures 276 and 277 represent the set of net planes having indices $h=3$, $k=2$, and $l=-2$. Figure 276 shows the traces of the

planes in the XY-plane; the planes themselves are *not* perpendicular to the XY-plane. However, their intercepts on the X- and Y-axes are in the ratio of the reciprocals of the first two of the Miller indices. Thus $(1/h)/(1/k) = (1/3)/(1/2) = 2/3$. Passing a plane through P_3 and the Z-axis we obtain Figure 277. This plane in which Figure 277 is drawn contains the normal to the net planes through the origin, ON, which passes through

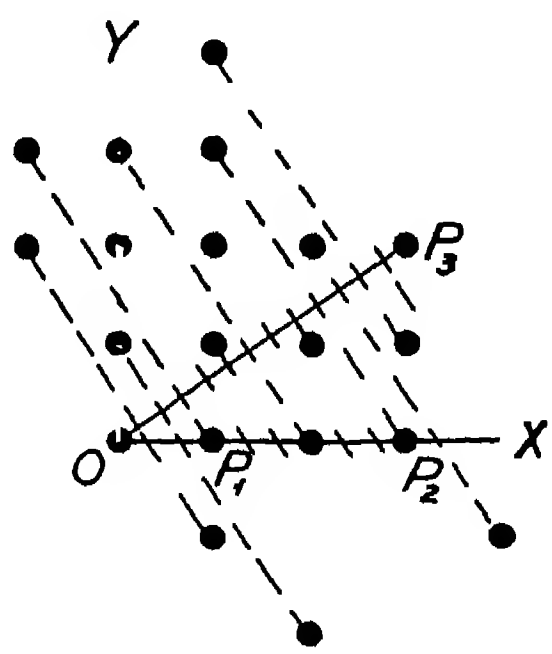


FIG. 276

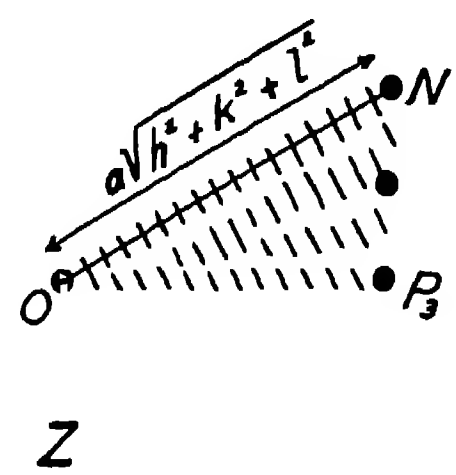


FIG. 277

the lattice point N. The distance from the origin to the first lattice point, N, through which the normal passes, is $a\sqrt{h^2+k^2+l^2}$—because the co-ordinates of N are (ha, ka, la), as above pointed out. It will presently be shown that there are net planes to the number of $(h^2+k^2+l^2)$ intercepting the length $a\sqrt{h^2+k^2+l^2}$ along the normal; the minimal distance between successive planes accordingly is:

$$D = \frac{a\sqrt{h^2+k^2+l^2}}{h^2+k^2+l^2} = \frac{a}{\sqrt{h^2+k^2+l^2}}. \tag{6}$$

The number of net planes $(h^2+k^2+l^2)$ is arrived at as follows: There are h of these planes cutting every line segment of length a parallel to the X-axis; this may be understood from Figure 276, for there are $h=3$ net planes corresponding to the distance $OP_1 = a$. Corresponding to the distance OP_2 $ha = 3a$, there are consequently $h^2 = 3^2$ net planes. Similarly, there are k planes cutting every line segment of length a parallel to the Y-axis. Co sponding to the distance $P_2P_3 = ka = 2a$, there are $k^2 = 2^2$ net planes. T are l net planes cutting every line segment of length a parallel to the Z-a and, corresponding to the distance P_3N (Fig. 277), there are $l^2 = (-2)^2$ planes. The total number of net planes corresponding to the distance is therefore $3^2+2^2+2^2 = 17$ in the illustration, or in the general case k^2+l^2, as above stated.

The Laue equations.—The Laue equations for the special case when the beam is incident along the positive Z-axis of a simple cubic lattice were derived in chapter 8, section 17. We shall now derive corresponding equations for the more general case in which the beam has any direction of incidence whatever.

Denote the direction angles of the incident beam by α_0, β_0, and γ_0, and the direction angles of the diffracted beam by α, β, and γ. Suppose in Figure 278 the circles represent a row of particles along one of the axes of coordinates—the X-axis, for example. The condition that disturbances from these particles shall reinforce each other is:

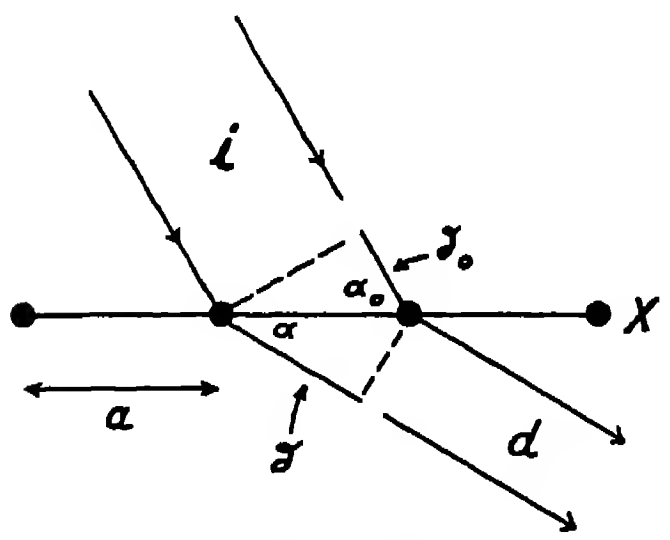


FIG. 278

$$(7) \qquad \delta-\delta_0=a\cos\alpha-a\cos\alpha_0=m_x\lambda\,.$$

The direction of diffraction generates a circular cone of angle α about the X-axis. The lower portion of the diagram is to be rotated about the row of particles as axis, the two halves of the diagram not being in general co-planar. The state of affairs is much the same as when the beam is incident along the Z-axis, but the angle of the cone will be different, and the cones for corresponding positive and negative orders will no longer be mirror images of each other in the YZ-plane. The previous state of symmetry has disappeared.

For rows of scattering particles along the Y- and Z-axes, figures similar to Figure 278 may be drawn and equations similar to equation (7) may be arrived at. The Z-axis no longer bears any especial distinction. The three fundamental equations thus become:

$$(8) \qquad \begin{aligned} m_x\lambda&=a\,(\cos\alpha-\cos\alpha_0)\,,\\ m_y\lambda&=a\,(\cos\beta-\cos\beta_0)\,,\\ m_z\lambda&=a\,(\cos\gamma-\cos\gamma_0)\,. \end{aligned}$$

In order that a principal maximum of diffraction may exist, the three cones about the three axes must all intersect in one line. For the direction of incidence the following equation applies:

$$\cos^2\alpha_0+\cos^2\beta_0+\cos^2\gamma_0=1\,.$$

For the direction of diffraction the corresponding equation holds. By forming the values of $\cos\alpha$, $\cos\beta$, and $\cos\gamma$ from equations (8), and squaring and adding these values, one finds an expression for the wave-length. This

expression is related to the one previously found in chapter 8, equation (7), but is more general, applying for any direction of incidence. The expression is:

$$\lambda = -2a\,\frac{m_x \cos \alpha_0 + m_y \cos \beta_0 + m_z \cos \gamma_0}{m_x^2 + m_y^2 + m_z^2}\,. \tag{10}$$

Equivalence of Bragg's and Laue's theories.—Denote by θ the angle made by the incident beam i (Fig. 279) with the plane OB which bisects the angle between the incident beam and the diffracted beam d and passes through the lattice point O. The angle between i and d is 2θ, and by a familiar geometric relation:

$$\cos 2\theta = \cos \alpha \cos \alpha_0 + \cos \beta \cos \beta_0 + \cos \gamma \cos \gamma_0\,. \tag{11}$$

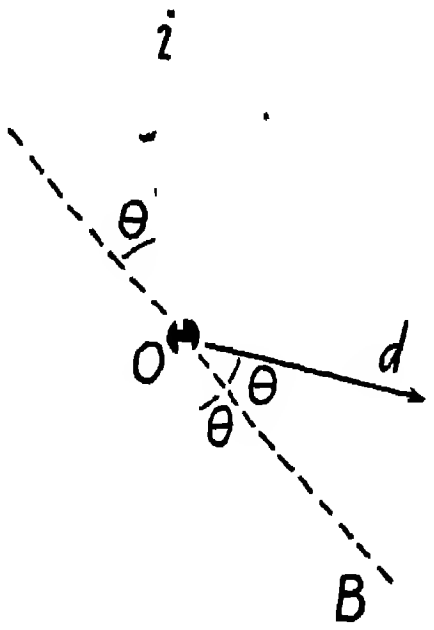


FIG. 279

By squaring and adding equations (8):

$$(m_x^2 + m_y^2 + m_z^2)\lambda^2 = a^2[(\cos \alpha - \cos \alpha_0)^2 + (\cos \beta - \cos \beta_0)^2 + (\cos \gamma - \cos \gamma_0)^2]\,. \tag{12}$$

Making use of equation (9) and the similar equation (5) of chapter 8 and also of equation (11), we have:

$$(m_x^2 + m_y^2 + m_z^2)\lambda^2 = 2a^2(1 - \cos 2\theta) = 4a^2 \sin^2 \theta\,. \tag{13}$$

Extracting the square root:

$$\sqrt{m_x^2 + m_y^2 + m_z^2} \cdot \lambda = 2a \sin \theta\,. \tag{14}$$

If m_x, m_y, and m_z have a common divisor m, we may write:

$$m_x = mM_x \qquad m_y = mM_y \qquad m_z = mM_z \tag{15}$$

and equation (14) becomes:

$$m\lambda = 2\,\frac{a}{\sqrt{M_x^2 + M_y^2 + M_z^2}} \sin \theta\,. \tag{16}$$

The numbers M_x, M_y, and M_z are integers which have no common divisor, or are prime to one another. It will now be shown that λ are the

Miller indices of a set of net planes which bisect the angle between the incident and diffracted beams: Referring to Figure 280, OB is the bisecting plane which passes through the lattice point O lying at the origin of co-ordinates. Let e denote another point in the plane OB. The broken lines ab and cd represent wave fronts in the incident and diffracted beams, respectively. The lengths of the optical paths aOc and bed are equal since OB is a bisecting plane. Conversely, when various paths between two wave fronts by way of a plane are all equal, this equality of paths insures that the plane bisects the angle between the incident and diffracted beams. Denote the length of the path aOc by Δ. Then the length of path between the wave fronts ab and cd by way of the nth lattice point on the positive X-axis (not shown in the figure) is: $\Delta - n \cdot m_x\lambda$, for in considering the individual lattice particles along the X-axis in succession the length of optical path by way of these particles is each time shortened by $m_x\lambda$, or $mM_x\lambda$. Now let us take $n = M_yM_z$. Then:

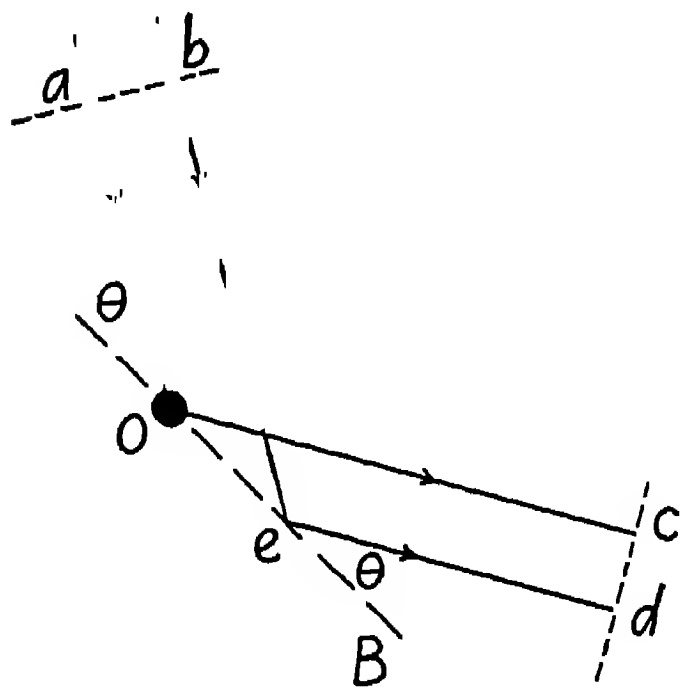


FIG. 280

$$\Delta - n \cdot m_x\lambda = \Delta - (M_yM_z)mM_x\lambda \equiv \Delta' . \tag{17}$$

Similarly, the length of path between the same wave fronts ab and cd by way of the n'th lattice point on the Y-axis, when $n' = M_zM_x$ is $\Delta - (M_zM_x) \cdot mM_y\lambda = \Delta'$. By way of the n''th lattice point on the Z-axis, when $n'' = M_xM_y$, the length of path is $\Delta - (M_xM_y)mM_z\lambda = \Delta'$. It is thus possible to find three lattice points, one on each of the three axes of co-ordinates, by way of which the length of the optical path between the wave fronts ab and cd is the same, being equal in each case to $\Delta - mM_xM_yM_z\lambda = \Delta'$. These three lattice points determine a plane which bisects the angle between the incident and diffracted beams. The plane must be occupied by other lattice points arrayed as shown in chapter 8, Figure 197, and must be a net plane. Furthermore, it may now be seen that OB must be a net plane, for it is parallel to the aforementioned plane, and, having one lattice point upon it, must be occupied by other lattice points in regular array. The two planes belong to a parallel set of net planes which bisect the angle between the incident and diffracted beams, and the planes of this set may hence be regarded as "reflecting" planes. Their intercept ratios are $n:n':n'' = M_y : M_zM_x : M_xM_y$ and their Miller indices are in the ratio of the reciprocals of the intercept ratios and are therefore (M_x, M_y, M_z). Introducing

these values of the indices into equation (6), we have for the minimal separation between successive net planes:

$$D = \frac{a}{\sqrt{M_x^2 + M_y^2 + M_z^2}} . \tag{18}$$

Thus equation (16) becomes:

$$m\lambda = 2\, D \sin \theta . \tag{19}$$

This will be recognized as Bragg's law. It has been deduced from the Laue equations. It is seen that m, the greatest common divisor of m_x, m_y, and m_z, represents the spectral order in the Bragg theory. The numbers m_x, m_y, and m_z, cleared of their common divisor, yield M_x, M_y, M_z, and (M_x, M_y, M_z) prove to be the Miller indices of a set of net planes which may be regarded as "reflecting" planes. The equivalence of the Bragg theory and the Laue theory is thus established.

APPENDIX H

THE REFRACTION OF ELECTRON WAVES

(Continuation of Chap. 9, Secs. 9 and 10)

We shall first deduce the directions of those Laue *X-ray* beams which are diffracted backward and fall into an A-azimuth of the nickel crystal, when the beam is incident normally. (These are also the directions of the corresponding electron beams *within* the crystal.)

Referring to Figure 281, consider the plane of the figure to be that of an A-azimuth (and B-azimuth). Let the small circles each represent the projection of a row of atoms into this plane. The orientation and separation of the reflecting net planes which give rise to the beams previously referred to as A, B, C, etc., are indicated in the present figure by the pairs of broken lines A, B, C, etc., respectively. Denote the angles which these reflecting net planes make with the horizontal by $\frac{1}{2}\theta_A'$, $\frac{1}{2}\theta'$, etc. These angles may be computed from the geometry of the crystal, for example, $\tan \frac{1}{2}\theta_A' = S/(16/3)a = 2.03/(16/3)\times 2.15 = .177$, or $\frac{1}{2}\theta_A' = 10°$. The colatitude angles of the X-ray beams are then θ_A', θ_B', etc. Computing in this manner, one obtains the data given in Table X. The values of $\sin \theta'$ of this table are the values of the abscissae for the X-ray beams of Figure 240 of chapter 9. The manner of evaluating these abscissae was there left to be explained later. Having now determined the abscissae, the values of the ordinates follow at once from the plane-grating relation. Or the ordinates may be deduced from Bragg's law after computing the separation of the net planes of each set, A, B, etc., from the geometry of the crystal; but this is more cumbersome.

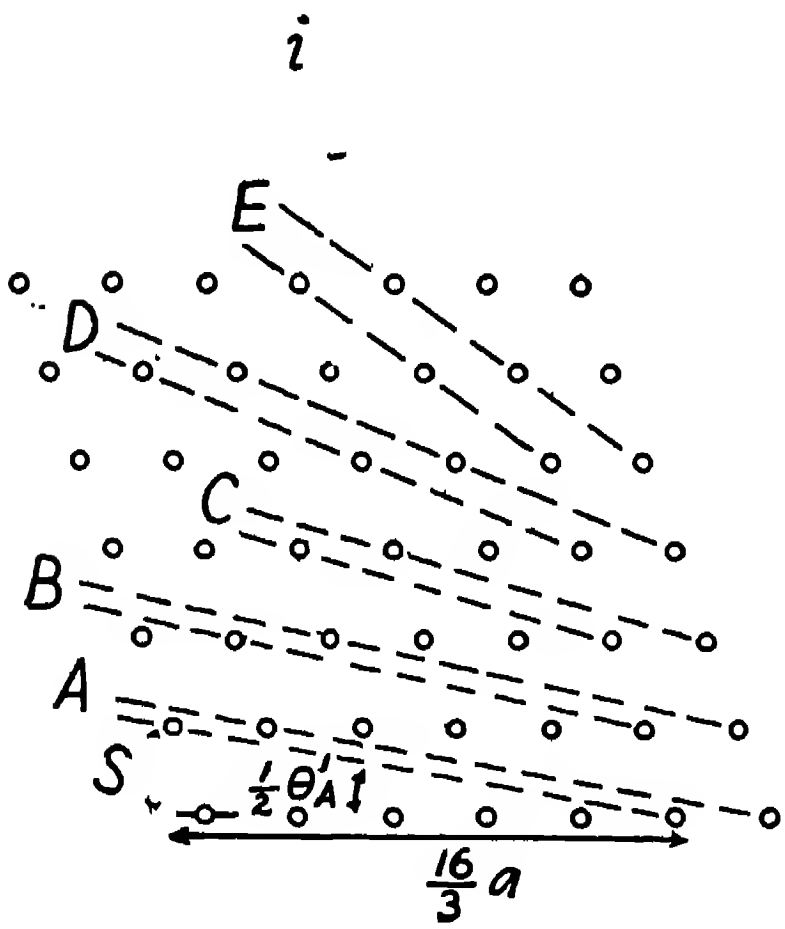


FIG. 281.—Reflecting net planes

Let us now see why the electron beams, when plotted as in Figure 240 of chapter 9, fall upon the same straight line through the origin as the X-ray beams: Referring to Figure 282, let the electron beam be incident from

above and let the solid circles each represent a row of atoms projected into a vertical plane which contains an A-azimuth. Consider the scattering of successive incident wave fronts by atoms of the second layer. This scattering will give rise to Huygens wavelets, and if the wave-length of the incident beam is the appropriate one, it will give rise to a Laue beam within the crystal which has the direction of the arrow D_x. Now the corresponding X-ray beam would have this same direction D_x and would continue in this direction after leaving the crystal, but for an electron beam the wave-length changes as the electrons pass into or out of the crystal, and, accordingly, the orthogonals to the wave fronts, the "rays," for the emerging beam, will be curved near the crystal surface, as illustrated. However, at some distance from the crystal the rays will again be straight and will have a direction D_e which makes an angle θ with the normal to the crystal surface such that $\sin\theta = \lambda/a$, for the first order ($m=1$). Accordingly, even though refraction occurs, the plane-grating relation is satisfied and the electron beam falls on the same straight line through the origin as the X-ray beam. Thus, taking account of various spectral orders, m, we have:

TABLE X

	θ'	$\sin\theta'$
A_x	20°0	.342
B_x	24 5	.415
C_x	31 6	524
D_x	44.0	.695
E_x	70.5	.943

$$\frac{\sin\theta}{\lambda} = \frac{m}{a} = \frac{\sin\theta'}{\lambda'} . \tag{1}$$

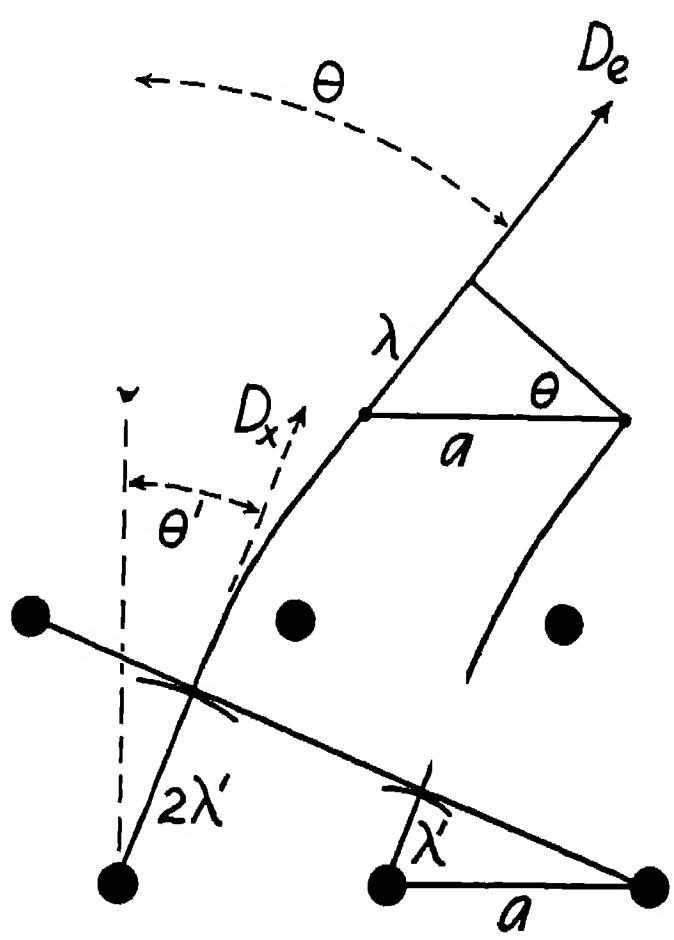


FIG. 282

The more general expression for Bragg's law, referred to in section 10 of chapter 9, which takes into account an index differing markedly from unity, will now be derived:

In Figure 283 let the arrow i, on the left, represent the incident beam, and let the broken line represent a wave front. In the lower part of the figure the first two layers of atoms of the crystal are represented by circles. The location of the atoms within these layers is really immaterial for our present purpose. A disturbance traveling to the second layer traverses a linear distance $2c$ within the crystal or an optical path $2\mu c$ within it. The

incident wave front will be reflected by the first layer of atoms to an instantaneous position W_1, and by the second layer to W_2. The retardation of W_2 relative to W_1, expressed in length of optical path, is:

$$\begin{aligned} \delta &= 2\mu c - a \\ &= \frac{2\mu D}{\cos\theta'} - 2\,b\sin\theta\,. \end{aligned} \tag{2}$$

But $b = D\tan\theta'$ and $\sin\theta = \mu\sin\theta'$, hence we arrive at the following equation, which is well known in the treatment of optical interference due to a plane-parallel plate, namely:

$$\delta = \frac{2\mu D}{\cos\theta'}(1-\sin^2\theta') = 2\mu\,D\cos\theta'\,. \tag{3}$$

The disturbance from the third layer of atoms is retarded by an equal amount with reference to that from the second layer, etc. In order to have

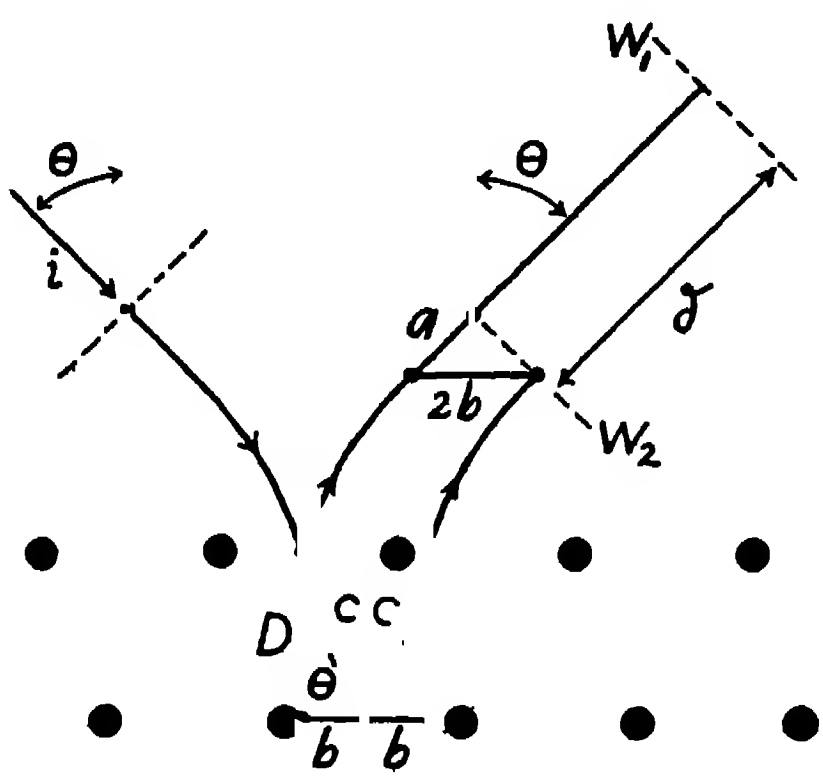


FIG. 283.—Bragg reflection of electron waves

constructive interference the path difference δ must equal a whole number of wave-lengths, $m\lambda$. Moreover, we may write

$$\cos\theta' = [1-\sin^2\theta']^{\frac{1}{2}} = \left[1-\frac{\sin^2\theta}{\mu^2}\right]^{\frac{1}{2}},$$

whence:

$$\delta = m\lambda = 2\mu D\cos\theta' = 2\mu\,D\left[1-\frac{\sin^2\theta}{\mu^2}\right]^{\frac{1}{2}},$$

or:

$$m\lambda = 2D\,[\mu^2-\sin^2\theta]^{\frac{1}{2}}\,, \tag{4}$$

which is the desired more general expression of Bragg's law and reduces to the law in its simple form when $\mu=1$. (The symbol θ here represents the angle of incidence, and *not* the glancing angle, whereas the law in its simple form is usually given in terms of the glancing angle.)

For X-rays Bragg's law can be used for determining wave-lengths. For electron waves the modified expression, equation (4), cannot be so used, for it involves μ, which is itself a function of the wave-length. The expression is useful, however, for other purposes. We have learned that electron wave-lengths are correctly given by the De Broglie relation, in particular by the equation $\lambda=12.25/V^{\frac{1}{2}}$. Inserting this value of λ in equation (4) and expressing the result in terms of μ as the unknown:

$$\mu=\left[\frac{150m^2}{4VD^2}+\sin^2\theta\right]^{\frac{1}{2}}. \tag{5}$$

From this equation the index of refraction can be calculated in terms of the bombarding potential, and also, by invoking equation (13) of chapter 9, evaluations of the inner potential can be made. These values, however, do not all turn out to be equal, and this fact reveals the insufficiency of the simple conception of the cause of the refraction of electron waves.

Finally, we shall show how the positions of the arrows in Figure 241 of chapter 9 were arrived at. The reader will recall that these positions indicate the bombarding potentials at which the beams would have developed if there had been no refraction. In the nickel crystal the distance D between successive layers of atoms parallel to a (1 1 1) face is 2.03 Å, and in the experiments in question θ was made equal to 10°. Hence applying Bragg's law in its simple form and then inserting for λ its equal $12.25/V^{\frac{1}{2}}$, we have:

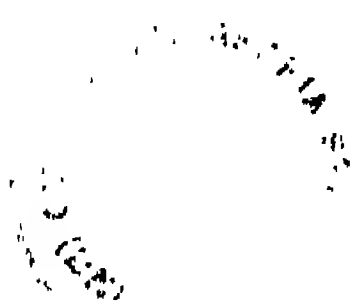
$$m\lambda=2\,D\cos\theta=2\times2.03\times\cos 10^\circ\,,$$

$$m\frac{12.25}{V^{\frac{1}{2}}}=2\times2.03\times.985\,,$$

or

$$V^{\frac{1}{2}}=\frac{12.25}{2\times2.03\times.985}m=3.06m\,.$$

Accordingly, the arrows occupy the positions $V^{\frac{1}{2}}=3.06$; 2×3.06; 3×3.06, etc.



INDEX

(The references are to pages; the letter *n* signifies footnote.)

Abbe, 211
  theory of microscopic vision, 212
Aberration, of concave grating, 152, 154 n., *430*
  permitted, 193
Abney mounting of grating, 162, 430
Abrupt error of ruling, of grating, 191
Absence of back wave, 266
Absent spectra, 133
Absorption spectra, of X-rays, 285
Accidental error of ruling, 190
Action, cross-section, of atoms, 395
  of grating, on single pulses, 255
  of prism, on single pulses, 259
Actual wave front, 85
Agitation, thermal, of atoms, 291, 296, 299
Airy, circular aperture, 90, *426*, 428
  theory of rainbow, 243
Alkali atoms, reflection of, 411
Allison, S. K., 348 n., 351, 358
  notation of, 357
Amorphous solids, diffraction of X-rays by, 359
Amplitude error, of ruling, 171
Analysis of crystals, methods of, 310
Anderson, J. A., elastic deformation of ruling machine, 173
  finishing of screw, 130 n.
  Lyman hosts, 184
  ruling of gratings, 128, 130
Andrade, X-ray spectrograph, 308, 338
Angle, Bragg, 331, 339
  glancing, 294
  of tolerance, 294, 338
Angular dispersion, of grating, 140
Aperture, circular, 31, 42, 77
  Airy's treatment, 426
  Fraunhofer pattern, 90
  graphical method, 83
  numerical, 208
  rectangular, 86, 96
  triangular, 107
Aplanatic surface, 200
Application, of Huygens' principle, 266
  of principle of interference, 268
Arago, 10, 20
  circular disk, 46
Areas of Fresnel zones, 30, 44
Arkadiew, 77, 81
Arnot, scattering of electrons, 395 n.
Astigmatism, concave grating, 157, 435
  caused by, curvature of grooves, 170
    error of run, 186
Asymptotic points, of Cornu spiral, 57
Atom, diffraction of, 399–411
  effective cross-section of, 395
  electron shells of, 283
  gun, 399
  structure factor, 291, 298
  thermal agitation of, 291, 296, 299
Atomic spectra, X-ray, 282, 284
Avogadro's number, 333 n.
Axis, zone, 328
Azimuth, curves, 372
  rotation in, 370

Babinet, 237
  principle of, 233
Back wave, absence of, 266
Bäcklin, 346, 348
Badger, laminary grating, 138 n.
Bailey, action cross-section, 396
Baily, concave grating, 157 n.
Bands, of Talbot, 253
Banerji, colors of mixed plates, 255 n.
Bar, pattern formed by, 3, 73
  spacing of internal fringes, 75
Barker, E. F., ruling of gratings, 128
Barkla, characteristic X-rays, 284
  polarization of X-rays, 231
Bartholinus, double refraction, 18
Basu, N., 277
Beams, electron, gas, 374
    plane-grating, 373
    space-lattice, 374
Bearden, X-rays, by grating, 346
Beets, value of *D* for calcite, 337
Berek, on Abbe's theory, 216
Bernal, rotation photographs, 313 n.
Bessel functions, 428
Bethe, inner or grating potential, 379
Beugungserscheinungen, Die, by Schwerd, 11.

Biot, 20
Blood corpuscles, diameter of, 240
Blue color of sky, 244, 250
Body-centered lattice, 290
Bouasse et Carrière, *Diffraction*, 78, 82, 254
Bowen, extreme ultra-violet, 168
  optimum width of grating, 170, 434
Box, Faraday, collector, 369
Bradley, velocity of light, 22
Bragg, W. H., evaluation of *D*, 336
  extinction of diffracted beam, 309
Bragg, W. H. and W. L., crystal analysis, 310
  focusing condition, 304
  X-ray spectrometer, 304
Bragg, W. L., on Kikuchi's experiment, 389
Bragg, W. L., James, and Bosanquet, 348 n., 358
Bragg angle, 331, 339
Bragg reflection, of electrons, 381, 458
  of X-rays, 292 ff.
Bragg wave-length, 331, 340
Bragg's law, 294
  correction of, 329
  generalized form, 458
Bragg's view, equivalence to Laue's, 453
Brewster, 20, 131
Bridge, theorems of, 107
Brockway and Pauling, 394 n.
Brode, scattering of electrons, 395 n.
Broese and Saayman, 396 n.
Broglie, L. de, 364
  phase waves, 365
  wave-length, 364, 368
Broglie, M. de, X-ray spectra, 305
Brougham, 5 n.
  antagonism of, toward Young, 10
Buchwald, error of ruling, 191
Buehl and Rupp, electron scattering, 396
Buisson, mounting of grating, 165 n.
Burns, mounting of grating, 165 n.

Cabannes, Rayleigh scattering, 245 n.
Calcite, value of *D* of Bragg's law, 336
Camera, pinhole, 80
Carrara, X-rays, by grating, 343
Cartwright, laminary grating, 138 n.
Cauchois, curved crystal, 306 n.
Cauchy, Fresnel integrals, 58 n.
Characteristic X-rays, 284
Chase, polarization, electron wave, 399
Chinmayan, 277
Circle, Rowland, 146
  with curved crystal, 306
Circular aperture, 31, 42, 77
  Airy's treatment, 426
  Fraunhofer pattern, 90
  graphical method for, 83
  and irregular obstacle, 51, 232
Circular obstacle, 45, 80
  graphical method for, 83
"Cis" form of dichloro-ethylene, 361, 393
Clark and Amber, 360 n.
Classification of X-rays, according to hardness, 284
  in series, 282
Coefficient of reflection, 354
Colatitude, 370
  curves, 372
Collector, Faraday, 369
Color of sky, 244, 250
Colors of mixed plates, 254
Common spectrometer, 139
Comparetti, 5 n.
Compensating for aberration by refocusing, 186, 194
Complementary screens, 233
Complete wave front, 43, 56
Compton, A. H., double crystal spectrometer, 348 n., 358
  grating at grazing incidence, 169
  polarization of X-rays, 281
  specular reflection of X-rays, 341
Compton, Beets, and De Foe, 337
Compton and Doan, 342
Concave grating, 127, *146*
  aberration of, 152, 154 n., *430*
  astigmatism, 157, 435
  for extreme-ultra-violet, 167, 434
  general focal relation, 155
  mountings for, 160
  numerical evaluations for, 152
  optimum width of, 170, 433, 434
  permitted aberration, 153
  permitted width, 433
  for soft X-rays, 170
Continuous X-ray spectrum, 282
Contribution from elementary area, 35
Coolidge, X-ray tube, 301
Cork, X-ray spectrograph, 309
  X-ray wave-lengths, by grating, 346
Cornu, 27, 59
  focal peculiarities of gratings, 190
  spiral, 56 ff.
    as vibration curve, 60, 415
Coronas, atmospheric, 240
Corpuscles, of blood, diameter, 240

Correction of Bragg's law, 329
Costa, Smyth, and Compton, 408 n.
Coster and Veldkamp, 286 n.
Crested fringes, 3
Critical angle, for X-rays, 341
Critique of Fresnel's theory, 264
Cross-section, effective, of atoms, 395
Crossed gratings, 132, 319
Crystal, analysis, methods of, 310
  curved, spectrograph, 306
  diffraction, of X-rays, 286 ff.
    of electrons, 362 ff.
    of atoms, 399 ff.
  liquid, 360
  of nickel, 370
Crystal, rotation of, 302
  rotation photographs, 311
Crystal, thin, diffraction of electrons by, 386
Crystal plane, 291
Crystal spectrometer, double, 348
Crystal structure, 288
Crystals, range of wave-lengths diffracted by, 299
Cubic lattice, 288
Curvature of grooves of grating, 170
Curve, azimuth, 372
  colatitude, 372
  dispersion, for electron waves, 380
  rocking, 351
  vibration, 38
  voltage, 372
Curved crystal spectrograph, 306
Cybotaxis, 359

Darwin, 309, 310 n., 335, 338
Datta, on Quetelet's rings, 244 n.
Dauvillier, 170
Davis, B., 348 n., 358
Davis, H. T., circular aperture, 79
Davisson and Germer, 362 ff.
  Bragg reflection of electrons, 381
  gas beams, 374 n.
  polarization of electron waves, 397
de Broglie; *see* Broglie
Debye, powder spectrograph, 315
  diffraction of X-rays by gases, 361
Definition, loss of, 172, 188
Deflected wave, loss of phase, 229
Deflection of light, 5
De Foe, 337
Deformation, elastic, of ruling machine, 173
De la Hire, 47 n.
Delisle, 5, 12 n., 47 n.
De lumine, by Grimaldi, 1, 3, 82
Dempster, 412
Depth and form of grooves, 125, 144
Dershem, reflection of X-rays, 344
Descartes, theory of rainbow, 243
Development of wave theory of light, 17
Diagram, vibration, 38
  of Laue spots, 327
Diameters, of particles, measurement of, 221, 238
  of stars, measurement of, 216
Dichloro-ethylene, 361, 393
Dieke, astigmatism of grating, 437, 442
Diffracted light and law of extreme path, 224
Diffracted wave front, 85 ff., 225 ff.
Diffracting edge, luminosity of, 46, *222* ff.
  observation close to, 264, 276
Diffracting system regarded simply as a barrier, 14, *264*, 270 ff.
Diffraction by apertures or obstacles of various types or by other diffracting systems, *see under type of diffracting system in question*
Diffraction, of atoms and molecules, 399–412
  dynamical theory, 26, 269
  of electrons, 362–99, 456
    by gases and liquids, 393
    by mica, Kikuchi's method, 386
    by ruled rating, 383
    by Thomson's method, 383
  Fraunhofer, 85 ff.
  Fresnel, 65 ff.
  kinematical theory of, 26
  at large angles, 275
  of material particles, 362
  of protons, 412
  and the quantum theory, 277
  of X-rays, 279 ff., 449
    by grating, 341
    Laue, 286, *315*, 452
    by liquids, amorphous solids and gases, 359
    by slit, 346
  Young's theory, 7, 230
Diffraction, Bouasse et Carrière, 78, 82, 254
Diffraction grating, 112
  action on single pulses, 255
  and Babinet's principle, 234
  for diffraction of, electrons, 383
    X-rays, 341
  resolving power of, 197
Diffraction patterns, virtual, 32, *104*

Diffraction phenomena, of Fraunhofer class, 85
of Fresnel class, 65
Diffusion rings, of Newton, 244
Dimensions of zones, 30, 44
Disk, circular, 45, 80, 83
Dispersion, of double crystal spectrometer, 357
of electron waves, 380
of grating, 114, 140
theory, applied to X-rays, 332
Dispersive power, 204
Distance between net planes in terms of Miller indices, 450
Distinctions between Fraunhofer and Fresnel diffraction, 65, *101*
Doan, X-rays, by grating, 342
Domini, De, theory of rainbow, 243
Double crystal spectrometer, 348
Double refraction of calcite, 18
Double slit, Fraunhofer pattern, 94, 116
Fresnel pattern, 76
Drucker, 360 n.
Dry objective, 207
Du Bois and Rubens, 133, 137 n.
Du Mond and Hoyt, 348 n., 358
Du Sejour, 5 n.
Dutour, 5 n.
Dymond, 395, 399
Dynamical theory of diffraction, 26, 269

Eagle, concave grating, astigmatism, 442
mounting for, 164, 433
Echelette, 144
Eckart, refraction of electrons, 379 n.
Edge, luminosity of, 46, *222* ff.
of geometrical shadow, 1, 66 ff., 421
observation close to, 276
straight, 2, *66*, 222, 419
Edges, absorption, for X-rays, 286
Edlén, optimum width of grating, 170, 434
Effect, Townsend-Ramsauer, 396
volume, with X-rays, 306
Effective cross-sections of atoms, 395
Ehrenberg and Mark, 348 n., 358
Ehrenberg and Susich, 348 n., 358
Einstein, 251, 366
Elastic deformation of ruling machine, 173
Elastic solid theory of ether, 21
Eldridge, velocity analyzer, 408 n.
Electromagnetic theory of light, advent of, 22
Electron beam, total reflection of, 379
Electron gun, 367
Electron shells of atom, 283
Electron waves, Bragg reflection, 381, 458
index of refraction for, 380, 459
polarization of, 396
refraction of, 377, 456
wave-length of, 364, 368
Electrons, diffraction of, 362–99
by grating, 383
by thin crystal, 386
Thomson's method, 383
Electrons, fast, 383
full speed, 369
scattering, by gases, 393
by liquids, 396
slow, 367
Elementary area, contribution from, 35
Elementary lune, 54, 415
Elementary zone, 38
Elements, periodic table of, 285
Ellett and Olson, 402
Ellett, Olson, and Zahl, 408 n., *410*
Elsasser, electron diffraction, 366
Equations, of Laue, 322, 452
Equidistant slits, 112, 116
Equivalence of Bragg's and Laue's theories, 300, 453
Eriometer, of Young, 238
Error of ruling of gratings, 170, 443
abrupt, 191
accidental, 190
curvature of grooves, 171
error of run, 185
error of spacing, 172, 443
lack of parallelism of grooves, 171
periodic, 172, 443
permissible periodic, 447
variation of groove form, 192
Estermann and Stern, 399 ff.
Etching of glass gratings, 169
Ether, elastic solid theory of, 21
Ewald, 309, 310 n., 338
Examples in Laue diffraction, 323
External fringes, 2, 70
attributed to interference, 9, 230
at square corner, 82
External total reflection, 170, 341
Extinction lines, of Kikuchi, 391
of Rutheford and Andrade, 309
Extreme path, law of, 199, 224
Extreme ultra-violet, 167, 434
Eyepiece, use of, for viewing fringes, 11

Fabry and Buisson, grating mounting, 165 n.
Face-centered lattice, 289
Factor, atom structure, 291, 298
Fagerberg, examination of grating, 180
False spectrum lines, 172, 174, 180, 185
Fan of rays, 435
Faraday, 23
  box collector, 369
Farnsworth, 374 n., *376*, 377 n., 382
Fast electrons, diffraction of, 383
  by thick single crystals, 391
Fermat, principle of least time, 199, 224
Fizeau, 22, 216
Focal peculiarities of gratings, 171, 186, 190
Focal relation, for concave grating, 155
Focusing condition, for X-rays, 303
Focusing along Rowland circle, 149
Form factor, 291, 298
Form of grooves of grating, 125
Formation of spectrum, by grating, 113
Foucault, 22, 193, 194
Fraunhofer, 21, *108*
  class of diffraction phenomena, 21, *85*
  diffraction, distinguishing features of, 101
    general principles of, 92
    optical system for, 85
  discovery of grating, 21, 111, 126
  Gesammelte Schriften, 111
  life and work of, 108
  lines, of solar spectrum, 21, 110
  pattern, circular aperture, 90, 426
    double slit, 94, 116
    rectangular aperture, 86, 96
  patterns, symmetry of, 106
  polarization, by grating, 136, 137 n.
Fresnel, 3, 5, 6, *10* ff.
  application of Huygens' principle, 15, 266
  class of diffraction patterns, 65
  critique of his theory, 264
  diffraction, distinguishing features, 101
  division of wave front into lunes, 417
  integrals, 58, 63, 418
  investigations and writings, 10, 13
  life of, 6
  Mémoire couroné, 12
  objections to Newton's views, 6
  Œuvres complètes, 17, 29 n., 417 n.
  principles of his theory, 14
  use of lens for viewing fringes, 11
  zones, 28
Friedrich, Knipping, and Laue, 286, 287 n., 315
Fringes, crested, 3
  external, 2, 9, 66, 74, 80
  hyperbolic loci of, 8, 69
  internal, 2, 9, 74, 80
  regarded as arising from interference, 9, 74, 230
Frisch, 399, 408
Full-speed electrons, 369

Gajewski, 361 n.
Gas beams, in electron diffraction, 374
Gases, scattering of electrons by, 393
  scattering of X-rays by, 359
Gauge, hot-wire, Pirani, 401
General principles of Fraunhofer diffraction, 92
Generalized form of Bragg's law, 458
Geometrical shadow, 2
  width of, 421
Geometrically illuminated region, 86, 104
Geometry of Cornu spiral, 56
Gerhardt, 221
Germer, 362 ff.
  Bragg reflection of electrons, 381
  gas beams, 374 n.
  polarization of electron waves, 397
Gesammelte Schriften, Fraunhofer, 111
Ghosts, formed by grating, 172
  Lyman, 180
  Rowland, 174, 443
    intensities of, 446
  very close to parent-line, 185
Gilbert, Fresnel integrals, 58 n.
Glancing angle, 294
Glaser, microphotographs, 126
Glasser, book concerning Roentgen, 279 n.
Glazebrook, on concave grating, 157 n.
Gleason, extreme ultra-violet, 170 n.
Glory, atmospheric, 242
Gouy, 267, 275
Graphical methods for circular aperture and obstacle, 83
Grating, 112
  action on single pulses, 255
  and Babinet's principle, 234
  concave, 127, *146*
    aberration, 430
    astigmatism, 157, 435
    mountings for, 160
  crossed, 13., 319
  depth and form of grooves, 125
  diffraction of electrons by, 383
  diffraction of X-rays by, 341
  discovery of, 111, 126
  dispersion of, 114, 140
  echelette, 144

error of ruling, 170, 443
false spectrum lines from, 172, 174, 180, 185
focal peculiarities of, 171, 186, 190
formation of spectrum, 113
grazing incidence, 169
having few lines, 116
having many lines, 120
ideal, 170
intensifying by etching, 169
laminary, 137
law, 113, 115
modified forms, 143, 343
machines for ruling, 128, 345
microphotographs of, 126
periodic structure as prime requisite, 124
plane, in practice, 138
polarization caused by, 136
reflection, 125
replicas, 130
resolving power of, 197
ruled, historical, 126
ruled through silver film, 132
spectra, intensity ratios, 133
spectrometer, 139
for infra-red, 142
spectrographs, 141, 160
wire, 132
Grating potential of crystal, 379
Grimaldi, discovery of diffraction, 1
crested fringes, 3
fringes at square corner, 82
Groove form, of grating, 125, 145
variation of, 192
Group period, 258
Group velocity, 256, 366
Gun, atom or molecule, 399, 409, 410
electron, 367
Gurney, reflection of positive ions, 414

Haga and Wind, 346
Hagenow, polarization of X-rays, 281
Half-period, lunes, 53
zones, 28 ff.
Halos, 237 ff.
Hamos, curved crystal, 306 n.
Hanawalt, absorption edges, 286 n.
Hancox, beams of metal atoms, 410
Hard X-rays, 284
spectrographs for, 306
Hardy, infra-red spectrometer, 144 n.
Heisenberg, wave mechanics, 366
Helium atoms, diffraction of, 403
Helmholtz, resolving power, 211
Hendricks, on Kikuchi's experiment, 390
Hertz, 23, 137
Hjalmar, correction of Bragg's law, 335
Hooke, 4
Hopkinson, discovery of grating, 5 n.
Hot-wire gauge, Pirani, 401
Hoyt, double spectrometer, 348 n., 358
Hufford, circular aperture, 79
Hughes, McMillen, and Webb, 395 n.
Hull, powder spectrograph, 315
Huygens, 15 ff., 266
Hydrogen, atomic, diffraction of, 408
molecular, diffraction of, 403
Hyperbolic loci of fringes, 9, 69

Ice crystals, atmospheric, 240
Ideal grating, 170
Illumination, normal, definition of, 43
Illustrative examples of Laue diffraction, 323
Immersion objective, 206
Incidence, grazing, with grating, 169
Inclination factor, 40
Inconsistency regarding phase, in Fresnel's theory, 44, 270, 273
Increment of path, 29
Index of refraction, for electrons, 380, 459
for X-rays, 293, 330, 334, 341, 343
Indices, Miller, 291, 449
and Laue order, 325, 328, 454
rational, law of, 449
Inflected wave, loss of phase, 225, 228
Inflection, 5
Inflexion, 4
Infra-red grating spectrometer, 142
Inner potential of crystal, 379, 382
Integrals, Fresnel, 58, 63, 418
Intensifying gratings by etching, 169
Intensities of ghosts, 175, 177, 446
Intensity ratios of grating spectra, 133
Intercept ratios, 291, 449
Internal fringes, 2, 80
attributed to interference, 9, 75
Ionization method, for X-rays, 304, 351
Irregular obstacle and circular aperture, 51
Ives, grating replicas, 131

Jentsch, diffraction at large angles, 275
Johann, curved cyrstal, 306 n.
Johannson, curved crystal, 306 n.
Johnson, B. K., microscope, 210 n.
Johnson, T. H., diffraction of atoms, 402, 408
Jordan, 5 n.
Jordan and Brode, 395 n.

Kalaschnikow, 275
Kayser, concave grating, 149 n.
Kellström, X-rays, by slit, 348
Kelvin, elastic solid theory, 21
Kerschbaum, atomic hydrogen, 409
Kiess, on Lyman ghosts, 184
Kievit and Lindsay, absorption edges, 286 n.
Kikuchi, electron diffraction, 386
Kinematical theory of diffraction, 26
Kirchhoff, dynamical theory, 271
Kirchner, electron diffraction, 382 n., 393 n.
  evaporation on to celluloid, 386
  on Kikuchi's experiment, 389
Kirkpatrick and Ross, 348 n.
Knauer and Stern, 399, 402
Knipping, 286, 315
Knochenhauer, Fresnel integrals, 58 n.
Knudsen, law of, 412
Koch, X-rays, by slit, 347
Kollath, 395 n., 396
Krishnamurti, 360 n.
Krypton, action cross-section, 396
Kulenkampff, 348 n., 358

Laboratory directions, for pattern formed by straight edge, 419
Lack of parallelism of grooves, 170
Lamina, thin, transparent, 252
Laminar zones, 417
Laminary grating, 137
Langmuir and Kingdon, 412
Langstroth, polarization, electron, 399 n.
Laplace, 20
Large angles, diffraction to, 275
Larsson, 333, 348
Lattice, body-centered, 290
  constant, for rocksalt, 289, 336
  constants for mica, 387
  face-centered, 289
  nickel, numerical data for, 372
  particles, row of, 316, 320
  potassium chloride, 288
  simple cubic, 290
  sodium chloride, 288
Laue, diagram, 327
  discovery of X-ray diffraction by crystals, 286
  equations, 322, 452
  on inner potential, 382 n.
  order, and Miller indices, 454
  spots, 287
  on Sugiura's experiment, 413
  theory, 315, 45.
    equivalence with Bragg's 453
Law of, extreme path, 199, 224
  gratings, 113, 115
    modified forms, 143, 343
  rational indices, 449
Least time, principle of, 199, 224
Le Cat, 5 n.
Lens, permitted aberration, 194
  use of, for viewing fringes, 11
Leonardo da Vinci, 4
Life of, Fraunhofer, 108
  Fresnel, 16
  Young, 7 n.
Lindsay, absorption edges, 286 n.
Linear dispersion, of grating, 140
Linnik, 348, 389
Liquid crystals, 360
Liquids, scattering of electrons by, 398
  scattering of X-rays by, 359
Lithium fluoride, 403
Littrow spectrograph, 142
Lloyd's mirror, 277, 348
Lommel, circular aperture, 78, 428 n.
Longacre, reflection of ions, 414
Loss of phase, of deflected wave, 229
  of inflected wave, 225, 228
Luminosity of diffracting edge, 222 ff.
  of disk or sphere, 46
Lummer, 99, 211 n.
Lunes, method of, 53, 415
  equivalent height of, 416
Lycopodium, diameter of particles, 240
Lyman, etching of gratings, 169 n.
  extreme ultra-violet, 167
  ghosts, 180

Machines for ruling gratings, 128
Mack, Stehn, and Edlén, 170, 434
McMillen, scattering of electrons, 395 n.
Maey, revival of Young's theory, 273 n., 275
Mairan, 5
Malus, polarization by reflection, 20
Mannkopf, astigmatism, 442
Many lines, gratings of, 120
Mascart, concave grating, 157
Massey and Mohr, 395 n.
Material particles, diffraction of, 362
Matter, wave theory of, 364
Maxwell, Clerk, 23
Maxwell, L. R., electron scattering, 396

Measurement of, diameters of particles, 221, 239
star diameters, 216
Meggers and Burns, concave grating, 165 n.
Meggers and Kiess, on Lyman ghosts, 184
Meibom and Rupp, protons, 413
Mémoire couronné, of Fresnel, 12
Metal atoms, reflection of, 410
Metals, positive ions of, reflection, 414
Meteorological phenomena, 240
Method of, lunes, 53, 415
rotating crystal, 302
Thomson, of electron diffraction, 383
zones, 28
Methods of crystal analysis, 310
Mica, diffraction of electrons by, 386
lattice constants for, 387
Michelson, abrupt error of grating, 191
ghosts formed by grating, 177
microscopic diameters, 221
star diameters, 216
Microphotographs of gratings, 126, 345
Microscope, resolving power of, 205, 212
Microscopic particles, measurement, 221
Microscopic vision, Abbe's theory, 212
Miller indices, 291, 449
and Laue order, 328, 454
Millikan, extreme ultra-violet, 168
Minkowski and Sponer, 396 n.
Mirror, permitted aberration, 194
Mitchell, astigmatism, 442
Mixed plates, colors of, 254
Modified forms of grating law, 143, 343
Mohr, scattering of electrons, 395 n.
Molecule gun, 399
Molecules, action cross-section of, 395
diffraction of, 399–408
Molybdenum trioxide coated plate, 409
Moore, on Abbe's theory, 216
Morse, inner potential, 382 n.
Moseley, law of, 283
method of diverging beam, 302
Mott, polarization of electron beam, 397
scattering of electrons, 395 n.
Mountings for concave grating, 160
Abney, 162, 430
Eagle, 164, 433
extreme ultra-violet, 167
parallel light, or stigmatic, 165
Paschen, 161
Rowland, 163, 430
Mueller, X-rays, organic crystals, 298 n.
Natural widths, of X-ray lines, 351
Nature of X-rays, 280
Net, 316, 318
Net plane, 291
Newton, diffusion rings, 244
on polarization, 19
views concerning diffraction, 5
Nickel crystal, 370
numerical data concerning, 372
Nishikawa, electron diffraction, 386
Nobert, ruling of gratings, 127
Normal illumination, definition of, 43
Notation of Allison, 357
Numerical aperture of microscope, 208
Numerical evaluations for concave grating, 152
Objective, dry, 207
immersion, 206
Obliquity factor, 40
Observation close to diffracting edge, 265, 276
Obstacle, circular, 45, 80
graphical method for, 83
Oersted, electromagnetic discovery, 23
Œuvres complètes, of Fresnel, 13, 29, 417
Oil-immersion objective, 206
Olson, reflection of atoms, 402, 408, *410*
Optical system for Fraunhofer diffraction, 85
Optimum width of grating, 170, 434
Order of diffraction and Miller indices, 325, 328, 454
Osgood, concave grating, 167 n., 169, 346
Overlapping of spectra, 115, 144
Palmer, scattering of electrons, 395 n.
Parallel-light mounting, 165
Parmelee, Clark, and Badger, 360 n.
Parratt, double crystal spectrometer, 348 n.
Particles, material, diffraction of, 362
small, diffraction by, 237 ff.
Paschen mounting, concave grating, 161
Path increment, 29
Pattern formed by ———; *see under type of diffracting system in question*
Patterns, Fraunhofer class, 86 ff.
symmetry of, 106
Fresnel class, 65
Kikuchi, 387
Laue, 287, 327
virtual, 104
Pauling, scattering of electrons, 394 n.

Pease, star diameters, 216 n., 221
Peirce, on Rowland ghosts, 177
Percentage of reflection, of X-rays, 354
Periodic error of spacing, 172 ff., 443
  permissible, 447
Periodic structure as prime requisite for existence of grating, 124
Periodic table of the elements, 285
Permitted aberration, 193
  for concave grating, 153, 432
Pernter and Exner, 240 n.
Pfund, infra-red grating spectrometer, 144
Phase error, in ruling grating, 171
Phase inconsistency in Fresnel's theory, 44, 270, 273
Phase loss, deflected wave, 229
  inflected wave, 225, 228
Phase velocity, 258, 366
Phase waves of de Broglie, 365
Phenomena, of Fraunhofer class, 85, 112
  of Fresnel class, 65
Physico-mathesis de lumine, of Grimaldi, 1, 82
Pinhole camera, 80
Pirani hot-wire gauge, 401
Planck, quantum theory, 25
Plane grating, in practice, 138
  infra-red spectrometer, 142
  spectrographs, 141
Plane-grating electron beams, 373
Plane-grating relation for electron beams, 373, 457
Plane net, diffraction by, 316, 318
Plane of polarization, 19
  and Rayleigh scattering, 247
Planes, spacing of, in terms of Miller indices, 450
Plates, mixed, colors of, 254
Pohl, X-rays, by slit, 347
Points, asymptotic, of Cornu spiral, 57
Poisson, 20
  circular disk, 46
Polanyi, Schiebold and Weissenberg, 313 n.
Polarization, 18
  of electron beam, 396
  by grating, 136
  and Rayleigh scattering, 247
  of X-rays, 281
Polygon, vibration, 38
Porter, on Abbe's theory of vision, 215
Positive ions, diffraction or reflection of, 412, 414
Potassium chloride lattice, 288
Potential, inner or grating, 379, 382
Powder spectrograph for X-rays, 313
Priestly, on Grimaldi's experiments, 4 n.
Principle, Babinet's, 233
  Huygens', application of, 15, 266, 272
  of interference, 16, 268
  of least time, Fermat's, 199, 224
Principles, of Fraunhofer diffraction, 92, 101 ff.
  of Fresnel diffraction, 28 ff., 53 ff., 101 ff.
  of Fresnel's theory, 14, 264
Prins, X-rays, dynamical theory of, 309, 310 n., 341
  diffracted, by grating, 344
    by liquids, 360 n.
Prism, action on single pulses, 259
  refraction of X-rays by, 331
  resolving power of, 202
Propagation, rectilinear, 47, 70
Protons, diffraction of, 412
Pulse hypothesis, of X-rays, 280
Pulses, isolated, action of grating on, 255
  action of prism on, 259
Purks, double crystal spectrometer, 348 n.

Quantum, relation, and X-rays, 282
  theory, advent of, 25
    and diffraction, 277
Quet, Fresnel integrals, 58 n.
Quetelet, rings, 244
Quincke, ghosts formed by grating, 177
  grating replicas, 131
  intensity of grating spectra, 136 n.
  laminary grating, 137 n.

Rabinov, X-rays, by slit, 347
Raether, electron diffraction, 382 n., 393 n.
Rainbow, 243
Raman, and Banerji, and Rao, 255 n.
Raman and Datta, 244 n.
Raman and Ramanathan, 359 n.
Ramsauer, 395 n., 396
Randall, Rooksby, and Cooper, 360 n.
Range, or angle, of tolerance, 338
  of wave-lengths, diffracted by crystals, 299
    comprised in X-ray region, 284
Rational indices, law of, 449
Ratios, intercept, 291, 449
Rayleigh, grating replicas, 131
  intensity of grating spectra, 136 n.
  permitted aberration, 193
  phase reversal zone plate, 33

poliarization by gratings, 137 n.
resolving power of, microscope, 211
prism, 205
telescope, 194, 196 n.
scattering, 244
Read, reflection of positive ions, 414
Rectangular aperture, 86, 96
Rectilinear propagation, 47, 70
Reference circle, Rowland, 146
Reflecting planes, Miller indices of, and Laue order, 325, 328, 454
Reflection, Bragg, 292
of electron waves, 381, 458
coefficient of, for X-rays, 354
critical angle of, 341
of metal atoms, 410
percentage of, for X-rays, 354
of positive ions, 414
specular, of X-rays, 341
total, external, 170
Reflection grating, 125
Refocusing to compensate for error, 186
Refraction, of electron waves, 377, 456
of X-rays by a prism, 331
Refractive index, for electrons, 380, 459
for X-rays, 293, 330, 334, 341, 343
Regular plane net, 316, 318
Reiche, 99, 211 n.
Reid, electron diffraction, 383, 386
Replicas, of gratings, 130
Resolving power, grating, 197
microscope, 205, 212
prism, 202
spectroscopic, definition of, 198
telescope, 194
X-ray, with double spectrometer, 350
Retardation, 29
of phase, of deflected wave, 229
of inflected wave, 225, 228
Rings, diffusion, of Newton, 244
of Quetelet, 244
Rittenhouse, 5 n.
discovery of grating, 111 n., 126
Rocking curve, 351
Rocksalt, lattice, 288
value of D of Bragg's law, 336
Rod, pattern formed by, 3, 73
spacing of internal fringes, 75
Roentgen, discovery of X-rays, 279
Rohr, von, on work of Fraunhofer, 111
Römer, velocity of light, 22
Ross, double crystal spectrometer, 348 n.
Rotating crystal, method of, 302
Rotation in azimuth, 370
Rotation photographs, from crystals, 311
Row of particles, diffraction by, 316, 320
Rowland, 127
concave grating, 127, 146
focal peculiarities of gratings, 190
ghosts, formed by grating, 174, 443
intensities of, 175, 446
machines for ruling gratings, 128
mounting for concave grating, 163, 430
scale of wave-lengths, 215
Rowland circle, 146, 149
with curved crystal, 306
Rubens, wire gratings, 133
Rubinowicz, revival of Young's theory, 232 n.
Rudolph, examination of gratings, 180
Ruled grating, historical, 126
for X-rays, 341, 345
Ruling, error of, of grating, 170
Ruling machine, 128
elastic deformation of, 173
Siegbahn's, for gratings for X-rays, 345
Run, error of, of grating, 185
Runge, concave grating, 149
astigmatism of, 442
on Lyman ghosts, 184
Rupp, electron diffraction, 374 n., 382 n., 396
by grating, 383
diffraction of protons, 413
polarization of electron beam, 399
Rutherford and Andrade, extinction lines, 309
spectrograph, 308, 338
Rutherfurd, gratings ruled by, 127

Saayman, electron scattering, 396 n.
Sandstrom, curved crystal, 306
Sawyer, R. A., extreme ultra-violet, 168
Sawyer, R. B., reflection of ions, 414
Scattering, of electrons, by gases and liquids, 393
of light, by small particles, 237 ff.
Rayleigh, 244
of X-rays, by gases liquids and amorphous solids, 359
Scheiner and Hirayama, 107
Scherer, powder spectrograph, 315
Schiebold, rotation photographs, 312
Schrödinger, wave mechanics, 366
Schumann, extreme ultra-violet, 167
Schuster, 43, 264 n.
laminar zones, 417
Schwarzschild, K., dynamical theory, 275
Schwarzschild, M., double crystal spectrometer, 348 n.

Schwerd, *Die Beugungserscheinungen*, 111
Screen, slightly transparent, 251
Screens, complementary, 233
Seemann, X-ray spectrographs, 307
Selenyi, vibration plane, 248 n.
Series classification of X-rays, 282
Shadow, geometrical, 2
ascertaining location of edge of, 421
Shells, electron, of atom, 283
Shinohara, electron diffraction, 393 n.
Siegbahn, ruling machine, 345
spectrograph for soft X-rays, 305
X-ray wave-length standards, 337
Simple cubic lattice, 290
Simple harmonic disturbances, summation of, 36
Single crystal, diffraction of fast electrons by, 391
Single slit, Fraunhofer pattern, 86
Fresnel pattern, 70
Single straight edge, 2, 66
laboratory directions for study of pattern, 419
Sirks, astigmatism of concave grating, 160, 442
Sky, blue color of, 244, 250
Slightly transparent screen, 251
Slit, double, Fraunhofer pattern, 94, 116
Fresnel pattern, 76
single, Fraunhofer pattern, 86
Fresnel pattern, 70
tapered, 82
Slit spectrograph, Seemann, X-ray, 307
Slits, equidistant, diffraction by, 112
Slow electrons, diffraction of, 367
Small number of lines, grating of, 116
Small particles, diffraction by, 237 ff.
Smoluchowski, 251
Sodium chloride lattice, 288
value of D of Bragg's law for, 336
Soft X-rays, 284
spectrographs for, 300
Sommerfeld, dynamical theory, 274
luminosity of diffracting edge, 232 n.
pulse breadth of X-rays, 34
Soret, zone plate, 32
Space-lattice diffraction, 292 ff., 321 ff., 452 ff.
of electrons, 374 ff., 456
Spacing of net planes in terms of Miller indices, 450
Sparrow, theory of imperfect gratings, 190 ff.
Spectra, grating, absent, 133
overlapping of, 115, 147
X-ray, 283
Spectral width of X-ray lines, 351
Spectrographs, concave-grating, 160
crystal, for hard X-rays, 306
powder, 313
for soft X-rays, 300
curved crystal, 306
plane-grating, 141
Spectrometer, crystal, Bragg, 304
double, 348
plane-grating, 139
infra-red, 142
Spectroscopic resolving power, 198
Spectrum, continuous, X-ray, 282
formation of, by grating, 112, 147
Spectrum lines, false, 172, 174, 180, 185
Specular reflection of X-rays, 341
Speculum, alloy, 127
Spiral, of Cornu, 56, 60, 415
Sponer, electron scattering, 396 n.
Spots, Laue, 287
Sproul, electron diffraction, 377
Square corner, external fringes, 82
internal fringes, 3
Standard wave-lengths, X-ray, 337
Star diameters, measurement of, 216
Stehn, optimum width of grating, 170, 434
Stempel, double spectrometer, 348 n., 358
Stenström, refraction of X-rays, 335, 341
Stephan, star diameters, 216
Stephenson, X-ray wave-lengths, 309 n.
Stern, 399 ff.
diffraction of atoms and molecules, 403
velocity analyzer, 408
Stewart, X-rays, by liquids, 359
Stigmatic mounting, concave grating, 165
Stokes, dynamical theory, 270
pulse hypothesis of X-rays, 280
Straight edge, 2, 66
laboratory directions for study of pattern, 419
Stratico, 5 n.
Straubel, symmetry of Fraunhofer patterns, 107
Structure of crystals, 288
Structure factor, 291, 298
Strutt; *see* Rayleigh
Sugiura, diffraction of protons, 412
Surface, aplanatic, 200
Susich, double spectrometer, 348 n., 358
Sylvite lattice, 288

Symmetry of Fraunhofer patterns, 106
Szilard, polarization, electron, 399

Table, periodic, of the elements, 285
Talbot's bands, 253
Tapered slit, 82
Tate and Palmer, 395 n.
Taylor, reflection of alkali atoms, 411
Telescope, resolving power of, 194
Temperature agitation of atoms, 291, 296, 299
Terrill, double spectrometer, 348 n.
Theorems of Bridge, 107
Theory of concave grating, 146
Theory, dynamical, 26, *269*
  Kirchhoff's, 271
  Sommerfeld's, 274
  Stokes's, 270
Theory, Fresnel's, 14, 28 ff., 53 ff., 264 ff.
Theory of microscopic vision, Abbe, 212
Theory, Young's, 7, 230
Thibaud, X-rays, by grating, 343, 346
Thin crystals, diffraction of electrons by, 386
Thin transparent lamina, 252
Thomson, G. P., electron diffraction, 377 n., 382 n., 393 n.
  method of, 383
    applied to, gases, 393
      liquid films, 396
    applied for diffraction of protons, 413
Thomson, J. J., pulse hypothesis, of X-rays, 280
  reflecting particle, 251
Thorp, grating replicas, 131
Tolerance, angle of, 294, 338
Total Bragg reflection of X-rays, 338
Total reflection, external, 170, 341
  internal, of electron beam, 379
Townsend-Ramsauer effect, 396
Traité de la lumière, Huygens, 15, 18
"Trans" form of dichloro-ethylene, 361, 393
Transmission grating, 112, 130 ff.
Trowbridge, echelette grating, 144
Tubes, X-ray, 301
Tyndall, scattering, 249, 250
Typical Fresnel patterns, 65

Ultra-violet, extreme, mountings of concave grating for, 167
Unit cell, of crystal lattice, 288

Van Lear, irregular obstacle and circular aperture, 52 n.
Variation of groove form, 171, 192
Veldkamp, absorption edges, 286 n.
Velocity, group, 256, 258, 366
  wave, or phase, 258, 366
Velocity analyzer, 408
Verdet, 47 n.
Vibration curve, 38 ff.
  Cornu spiral as, 60
Vibration diagram, 38
Vibration plane, and Rayleigh scattering, 247
Vibration polygon, 38
  based upon elementary lunes, 55
Views of, Newton, 5
  Young, 7
Vinci, Leonardo da, 4
Virtual diffraction patterns, 104
Voltage curves, electron diffraction, 372
Volume effect, 306
Vorlaender, 360 n.

Wadsworth, grating mounting, 165 n., 167
Wagner and Kulenkampff, 348 n., 358
Wallace, grating replicas, 131
Walter and Pohl, 347
Watson, scattering of electrons, 395
Wave, deflected, loss of phase, 229
  inflected, loss of phase, 225, 228
Wave front, actual, 85
  complete, representation of effect of, 43 56
  diffracted, 85
Wave group, 256
Wave-length, Bragg, 331, 340
  de Broglie, electron, 364, 368
  range, of X-ray region, 284
    diffracted by crystals, 299
  scale, of Rowland, 115
Wave-lengths, standard, X-ray, 337
Wave mechanics, 366
Wave motion, equations of, 34
Wave theory, of light, development of, 17
  of matter, 364
Wave velocity, 258, 366
Waves, electron, Bragg reflection of, 381, 458
    dispersion curve, 380
    index of refraction, 380, 459
    polarization, 396
    refraction of, 377, 456
    wave-length of, 364, 368
Webb, scattering of electrons, 395 n.

Wedge spectrograph, Seemann, 307
Whewell, 20 n., 244
White radiation, 315, 365
Width, permitted, or optimum, of concave grating, 153, 170, 433, 434
Wien, diffraction at large angles, 275
Wierl, electron diffraction, 386, 394
Williams, double crystal spectrometer, 348 n., 351, 357
Wind, X-rays, by slit, 346
Wire grating, 132
Wollan, 298 n., 310 n., 361 n.
Wollaston, solar lines, 110 n.
Wood, R. W., discharge tube, 409
  echelette grating, 144
  examining gratings, 178 n., 180
  ghosts close to parent-line, 185
  intensifying gratings by etching, 169
  on Lyman ghosts, 184
  phase reversal zone plates, 33
  polarization by gratings, 137 n.
  ruling of gratings, 128
Worsnop, electrons, by grating, 383
Wyckoff, 310 n.

X-ray absorption spectra, 285
X-rays, characteristic, 284
  classification, according to hardness, 284
    in series, 282
  coefficient of reflection, 354
  continuous spectrum, 282
  critical angle of reflection, 341
  diffraction, by gratings, 341
    by liquids, amorphous solids and gases, 359
    by slit, 346
  discovery of, 279
  effects of absorption, 341, 344
  focusing condition, 303
  gratings for, 345
  hard, spectrographs for, 306
  index of refraction, 293, 330, 334, 341, 343
  ionization method, 304
  nature of, 280
    spectra, 283
  percentage of reflection, 354
  polarization, 281
  powder, spectrograph, 313
  pulse hypothesis, 230
  quantum relation for, 282
  refraction by prism, 331
  series classification of, 282
  soft, spectrographs for, 300
  specular reflection, 341
  standard wave-lengths, 337
  total Bragg reflection, 338
  tubes, Coolidge, and gas, 301
  very soft, 284
    investigation of spectra, 299, 346
X-unit, definition of, 284

Yamaguti, electron diffraction, 393 n.
Young, colors of mixed plates, 254
  eriometer, 238
  life of, 7 n.
  theory of diffraction, 7, 230

Zabel, diffraction of atoms, 408
Zahl, 408 n., 410
Zone axis, 328
Zone plates, 32
Zones, dimensions of, 30, 44
  elementary, 38
  Fresnel, 23 ff., 77 ff.
  laminar, 417
  method of, 28

Acc. No. 2144
Class No. B. 9.
Book No. 30

PRINTED IN USA